DOCUMENT RESUME

ED 052 653 FL 002 383

TITLE	Speech Research: A Report on the Status and Progress of Studies on the Nature of Speech, Instrumentation for Its Investigation, and Practical Applications. 1 October - 31 December 1970.
INSTITUTION	Haskins Labs., New Haven, Conn.
SPONS AGENCY	Office of Naval Research, Washington, D.C. Information Systems Research.
REPORT NO	SR-24
PUB DATE	Jan 71
NOTE	184p.
EDRS PRICE	EDRS Price MF-$0.65 HC-$6.58
DESCRIPTORS	Animal Behavior, *Articulation (Speech), Artificial Speech, Behavioral Science Research, Computer Programs, Cross Cultural Studies, Dyslexia, English, *Laboratory Experiments, *Language Research, Letters (Alphabet), Linguistic Performance, Phonetics, *Psycholinguistics, Reading Development, Reading Difficulty, Russian, *Speech, Speech Pathology

ABSTRACT

This report is one of a regular series on the status and progress of studies on the nature of speech, instrumentation for its investigation, and practical implications. Manuscripts and extended reports cover the following topics: (1) "On Learning a New Contrast," (2) "Letter Confusions and reversals of Sequence in the Beginning Reader: Implications for Orton's Theory of Developmental Dyslexia," (3) "Perception of Dichotically Presented Steady-State Vowels as a Function of Interaural Delay," (4) "Perceptual Competition Between Speech and Nonspeech," (5) "Temporal Order Perception of a Reversible Phoneme Cluster," (6) "Phonetic Ability and Related Anatomy of the Newborn and Adult Human, Neanderthal Man, and the Chimpanzee," (7) "A Voice for the Laboratory Computer," (8) "Audible Outputs of Reading Machines for the Blind," and (9) "THESIS: Temporal Factors in Perception of Dichotically Presented Stop Consonants and Vowels." (Author)

ED052653

SR-24 (1970)

SPEECH RESEARCH

A Report on
the Status and Progress of Studies on
the Nature of Speech, Instrumentation
for its Investigation, and Practical
Applications

1 October - 31 December 1970

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

Haskins Laboratories
270 Crown Street
New Haven, Conn. 06510

Distribution of this document is unlimited.

(This document contains no information not freely available to the general public. Haskins Laboratories distributes it primarily for library use.)

FL002383

ACKNOWLEDGMENTS

The research reported here was made possible in part by support from the following sources:

Information Systems Branch, Office of Naval Research
Contract N00014-67-A-0129-0001
Req. No. NR 048-225

National Institute of Dental Research
Grant DE-01774

National Institute of Child Health and Human Development
Grant HD-01994

Research and Development Division of the Prosthetic and Sensory Aids Service, Veteran Administration
Contract V-1005M-1253

National Institutes of Health
General Research Support Grant FR-5596

Connecticut Research Commission
Grant Award No. RSA-70-9

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)*
Haskins Laboratories, Inc.
270 Crown Street
New Haven, Conn. 06510

2a. REPORT SECURITY CLASSIFICATION
Unclassified

2b. GROUP
N/A

3. REPORT TITLE
Status Report on Speech Research, No. 24, October-December 1970

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*
Interim Scientific Report

5. AUTHOR(S) *(First name, middle initial, last name)*
Staff of Haskins Laboratories; Franklin S. Cooper, P.I.

6. REPORT DATE
January 1971

7a. TOTAL NO. OF PAGES
185

7b. NO. OF REFS
160

8a. CONTRACT OR GRANT NO.
ONR Contract N00014-67-A-0129-0001

b. PROJECT NO.
NIDR: Grant DE-01774
c. NICHD: Grant HD-01994
NIH/DRFR: Grant FR-5596
VA/PSAS Contract V-1005M-1253
d. Conn. Res. Com.: Grant RSA-70-9

9a. ORIGINATOR'S REPORT NUMBER(S)
SR-24 (1970)

9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*
None

10. DISTRIBUTION STATEMENT
Distribution of this document is unlimited.*

11. SUPPLEMENTARY NOTES
N/A

12. SPONSORING MILITARY ACTIVITY
See No. 8

13. ABSTRACT

This report (for 1 October - 31 December) is one of a regular series on the status and progress of studies on the nature of speech, instrumentation for its investigation, and practical implications. Manuscripts and extended reports cover the following topics:

On Learning a New Contrast

Letter Confusions and Reversals of Sequence in the Beginning Reader: Implications for Orton's Theory of Developmental Dyslexia

Perception of Dichotically Presented Steady-State Vowels as a Function of Interaural Delay

Perceptual Competition Between Speech and Nonspeech

Temporal Order Perception of a Reversible Phoneme Cluster

Phonetic Ability and Related Anatomy of the Newborn and Adult Human, Neanderthal Man, and the Chimpanzee

A Voice for the Laboratory Computer

Audible Outputs of Reading Machines for the Blind

THESIS: Temporal Factors in Perception of Dichotically Presented Stop Consonants and Vowels

DD FORM 1 NOV 68 1473 (PAGE 1)
S/N 0101-807-6811

*This document contains no information not freely available to the general public. Its distribution is primarily for library use.

UNCLASSIFIED
Security Classification

A-31408

UNCLASSIFIED
Security Classification

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Cross-language studies (Russian/English)						
Labelling behavior						
Developmental dyslexia						
Letter confusions and order reversals						
Speech perception						
Dichotic listening experiments						
Dichotic lag effect						
Dichotic speech/nonspeech						
Temporal order judgments						
Speech capabilities of primates						
Voice answer-back from computers						
Reading machines for the blind						
Speech synthesis by rule						
Compiled speech						

DD FORM 1 NOV 65 1473 (BACK)
S/N 0101-807-6821

UNCLASSIFIED
Security Classification

A-31409

CONTENTS

PART I. Manuscripts and Extended Reports

On Learning a New Contrast 1

Letter Confusions and Reversals of Sequence in the Beginning Reader: Implications for Orton's Theory of Developmental Dyslexia 17

Perception of Dichotically Presented Steady-State Vowels as a Function of Interaural Delay 31

Perceptual Competition Between Speech and Nonspeech 35

Temporal Order Perception of a Reversible Phoneme Cluster . 47

Phonetic Ability and Related Anatomy of the Newborn and Adult Human, Neanderthal Man, and the Chimpanzee 57

A Voice for the Laboratory Computer 91

Audible Outputs of Reading Machines for the Blind 97

PART II. Thesis

Temporal Factors in Perception of Dichotically Presented Stop Consonants and Vowels 1

PART III. Publications and Reports

PART I: MANUSCRIPTS

AND

EXTENDED REPORTS

On Learning a New Contrast

Leigh Lisker[+]
Haskins Laboratories, New Haven

Descriptions of the speech behavior of human beings, both their management of the vocal tract and their perceptual processing of its audible output, require data on a wide variety of talkers if we are to separate the biological and cultural factors which govern speech activity or, perhaps more realistically, if we are simply to distinguish between the features which characterize speech generally and those specific to particular kinds of speech. One obvious way of checking on the degree of universality of generalizations concerning speech behavior is to compare speakers of diverse languages. We would suppose, if they manage phoneme inventories that differ in size and distinctive phonetic properties, that this is no more to be connected with physical characteristics of the speech-producing and speech-perceiving mechanisms than are grammatical differences in the languages which their speech "implements." Phonetic differences between two languages presumably reflect either different choices from some general inventory of phonetic dimensions which are, in principle, equally available to all language users or different ways of exploiting the same phonetic dimensions. Thus, for example, the feature of glottalization may serve to differentiate consonants in one language and not in another, whose speakers nonetheless might readily distinguish, if they had to,[1] between glottalized and unglottalized consonants. In another case, two languages might make use of very much the same range of vowel sounds but differ as to just how these are grouped into categories. Here the problem for the speaker of one language learning the other would be to adopt new criteria for deciding which sounds were the same and which different.

In comparing the language behavior of speakers of diverse languages we may follow certain psycholinguistic testing procedures which involve speech or speech-like auditory stimuli. Differences in test performance may in general be taken to reflect differences in the subjects' linguistic backgrounds, i.e., they represent an effect of learning. What is not quite clear is exactly what it is that was learned or the extent to which this learning may be said to have affected permanently the ability of speakers of one language to learn to match the performance of speakers of another. Over the past dozen years researchers at the Haskins Laboratories have been testing subjects, for the most part speakers of American English,

---

[+]Also, University of Pennsylvania, Philadelphia.

[1]As in a phonetic or psycholinguistic exercise, for example. If subjects discriminate between items they call "the same" so far as differentiating words of their own language, then this counters the view that linguistic coding always intervenes between the peripheral processing of the acoustic signal and the execution of the discrimination task.

to determine a relation between their linguistic identification of synthetic speech stimuli and their ability to detect the acoustic differences between individual test stimuli. In tests involving the presentation of steady-state synthetic vowel sounds (Fry et al., 1962; Stevens et al., 1969) no very close connection was found between subjects' identification of stimuli with English vowels and their ability to distinguish them in a conventional ABX test; items labelled alike were almost as easy to discriminate in this test as were items labelled differently. On the other hand, for a stimulus set whose members were distributed among the categories ba, da, and ga, it appeared that subjects were able to distinguish only those items which they assigned to different categories. Thus the results of testing for isolated vowel and initial stop discrimination appeared to be radically different.[2]

Unfortunately, in the testing program just referred to, no very extensive cross-language data have as yet been collected. However, what has so far been done in this area does not disconfirm the notion of a quite different relationship between labelling and discrimination behavior for vowels and stop consonants. In one vowel study (Stevens et al., 1969) American and Swedish subjects were compared in respect to the labelling and discrimination of certain vowel-like stimuli having very different categorial status in English and Swedish. It was found that linguistic experience, as reflected in differences in the way in which Swedes and Americans classified the test stimuli, had little apparent effect on their ability to discriminate. A cross-language consonant study (Abramson and Lisker, 1970; Lisker and Abramson, 1970), of which the present paper is a continuation, involved the comparison of several groups of speakers with respect to the dimension of voicing as this serves to distinguish between categories of stops in synthesized consonant-vowel syllables. Comparison of the labelling and discrimination behavior of groups of English-, Spanish-, and Thai-speaking subjects showed differences in the use of voicing as a feature, whereby stop categories in the three languages are phonetically distinguishable. It will be useful here to review briefly the background and findings of this study.

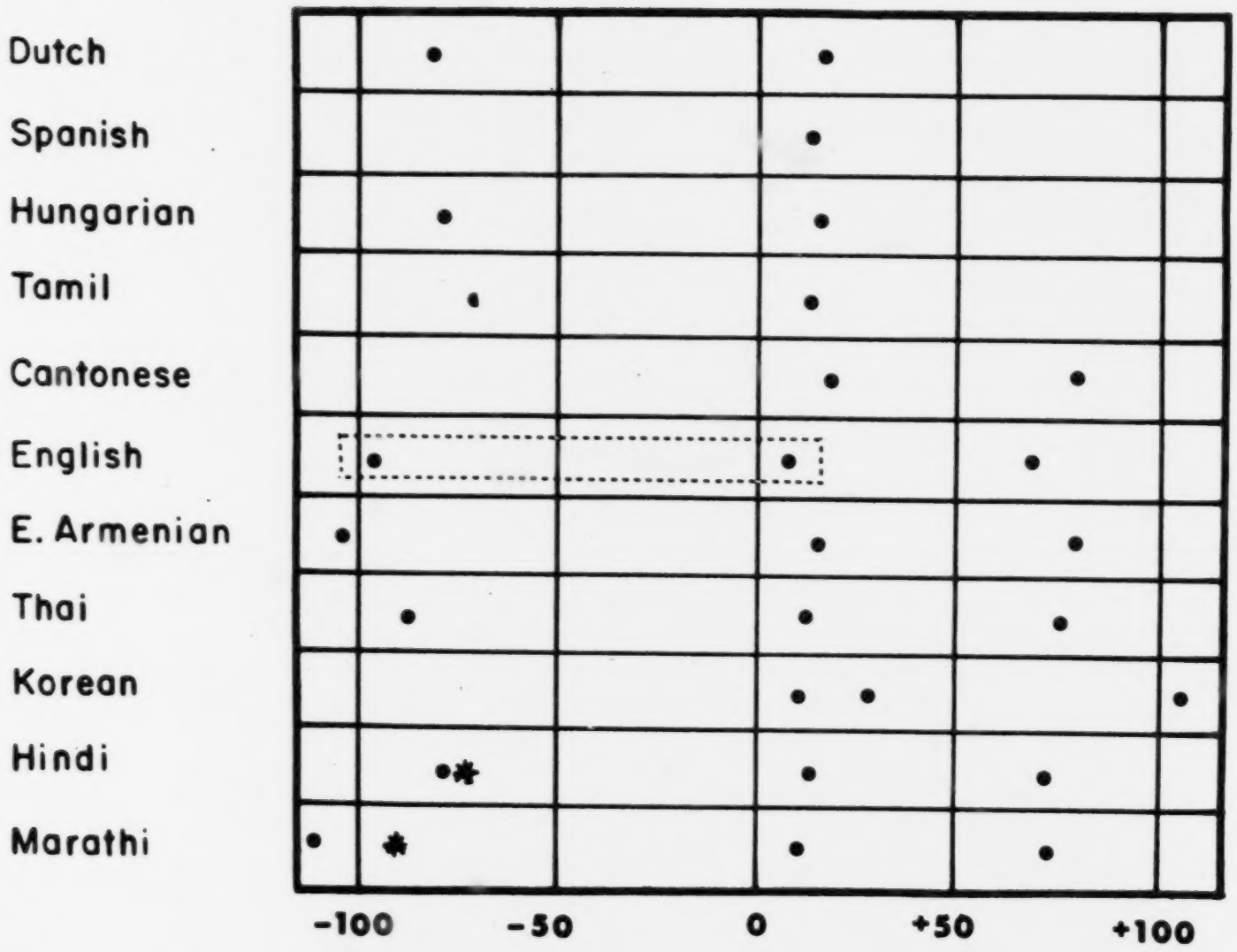
In very many of the world's languages, distinctive use is made of two, and sometimes three, categories of initial prevocalic stop consonants which differ, among other things, in the extent to which the larynx participates in their production, as this can be measured by determining the time of voice onset relative to that of release of the occlusion. Spectrographic examination of the word-initial stops in a number of languages (Lisker and Abramson, 1964) has shown that there are significant differences in voice onset timing ("VOT") from language to language but that the placement

---

[2] Whether these differences can be taken as evidence for a motor theory of speech perception may be regarded as doubtful, since the acoustic variables involved in the vowel and stop-consonant studies differ markedly. Nor can one readily assume that conclusions based on testing of steady-state vowel patterns will be valid for the perception of vowels in running speech. At the same time, however, evidence that the perception of speech differs significantly from that of nonspeech auditory stimuli, where this difference cannot be attributed to purely acoustic factors, have been recently reported (see Mattingly et al., 1969).

of category boundaries along this dimension is hardly random. Measurement data (Fig. 1) derived from productions of isolated words in a dozen languages suggest that there are three preferred timing relations between voicing onset and stop release: voicing begins almost 100 msecs before release (VOT ≅ -90); it begins at or just after the release (VOT ≅ +10); or it begins well after the release (VOT ≅ +75). These values, we may assume, correspond respectively to the voiced, voiceless unaspirated, and voiceless aspirated stops of classical descriptive phonetics. For all but one of the languages in our sample, Korean, it can be said that they differ essentially in the number and selection of stops from this set of phonetic categories. Thus Dutch, Spanish, Tamil, and Hungarian each make use of the two categories at VOT ≅ -90 and VOT ≅ +10. Cantonese differs from these languages in that, while it too has two stop categories, they involve VOT values at about +10 and +75. Languages with three categories along the VOT dimension (Eastern Armenian, Thai, Hindi, and Marathi, but not Korean) simply select all three of the categories described. Two of the languages examined are anomolous, Korean and English. Korean is a three-category language, but in initial position all of its stops are voiceless, i.e., voicing begins only with or following release, with VOT values at roughly +10, +30, and +100. English is peculiar in being a two-category language that utilizes all three VOT values but does not distinguish initially between stops with voicing lead and those for which VOT ≅ +10. The choice between the two appears to be, for American speakers of English at least, a choice that is partly idiosyncratic (Lisker and Abramson, 1964: 395) and partly a matter of style (Lisker and Abramson, 1967: 20-24).

In our comparison of different languages with respect to the timing of voice onset, we have been talking as though the only differences among the stop categories being compared are those of voice onset timing. Acoustically, of course, this is very far from being the truth; the effect of a small change in the timing of voice onset is very different, depending on whether onset precedes, coincides with, or follows the stop release. What we have been calling the VOT continuum is, at best, a continuum only in the articulatory domain, provided we make the no doubt oversimplifying assumption that a series of syllables such as [ba, b̥a, pa, p'a, pʰa] can be produced by executing an articulatory program that is invariant in respect to those components which effect the supraglottal gestures and that differs only in respect to those which determine when the laryngeal signal begins. This assumption, which derives from Dudley's well-known model of vocal-tract operation (Dudley, 1940), is most unlikely to be justifiable in detail. However, it is true enough to be convenient, for it provides the rational for a relatively simple program for the synthesis of stop-vowel syllables. For a particular syllable in which a fixed spectral pattern determines, for example, that it will be identified as consisting of a labial stop and the vowel [a], the voicing state of the stop is controlled by specifying the time at which the signal "exciting" the pattern shifts from one of aperiodic to one of periodic type. Thus for each place of stop closure a series of syllables could be generated whose initial consonants ranged from fully voiced (with onset of pulsing well before the burst marking stop release) to heavily aspirated voiceless stops (with pulsing onset considerably after the burst). Spectrograms of sample syllables, produced by means of the Haskins Laboratories' parallel



Mean voicing-onset timings for stops in word-initial position (from Lisker and Abramson 1964). Stop release is at 0 on abscissa, which represents no. of milliseconds by which voicing onset precedes (negative values) or follows (positive values) release. Starred entries are for voiced aspirates.

Fig. 1

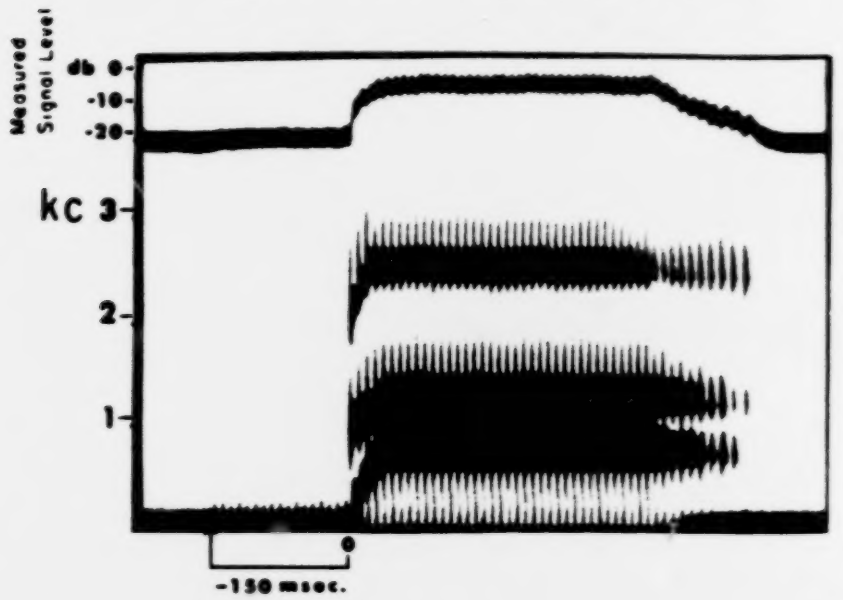
resonance synthesizer under computer control (Mattingly, 1968), are illustrated in Figure 2. The upper pattern illustrates the case in which the spectral pattern is excited from start to finish by a periodic signal. Four successive segments of the pattern may be distinguished: an initial segment characterized by a single, very low-frequency formant with a duration of 150 msecs, which corresponds to the articulatory closure; a burst or transient of about 10 msecs, which corresponds to the release of the stop; an interval of about 50 msecs with three formants of shifting frequencies, which "transition" corresponds to the interval of articulatory movements from closed stop to open vowel state of the oral cavity; a final segment with three formants at fixed frequencies for 450 msecs, corresponding to the vowel [a] as produced with "steady-state" articulation. The pattern immediately below represents the case where the periodic excitation begins just after the burst, while the lower pattern shows the synthesizer output where the same excitation begins 100 msecs after the burst. In both of the latter patterns the interval beginning with the burst and ending with the onset of pulsing is excited by an aperiodic signal. Particularly in the lowermost pattern, which listeners identify as a syllable beginning with a heavily aspirated stop [p^h], we can observe the feature of "first formant cutback," i.e., complete suppression of the first formant over the interval of noise excitation of the burst and upper formants. This feature must be considered not independent of the choice of excitation type (Liberman et al., 1958). We may suppose that the association between this spectral difference and the voicing dimension arises because the larynx is not only a signal source, but because, as a part of the cavity system of the vocal tract, its state helps determine the resonance properties of the tract. One might then associate with the absence of pulsing a large attenuation of the first formant, provided a fairly large glottal aperture during the interval between release and voicing onset is assumed.[3] Equally well, perhaps, one might suppose that the transmission characteristics of the tract are essentially identical for periodic and aperiodic source signals, but that the aperiodic source is deficient in intensity over the frequency range below the second formant. In any case a close relation both in production and in perception, between the onset of pulsing and the fairly rapid development of first-formant intensity to a level appropriate to the following vowel was established from spectrographic study of real speech and on the basis of preliminary perception testing of synthetic stop-vowel patterns in which the VOT and first-formant cutback features were independently manipulated. Consequently, in our further discussion of what we have been calling the "VOT dimension," it is to be understood that the onset of pulsing is regularly accompanied by the simultaneous onset of the first formant.

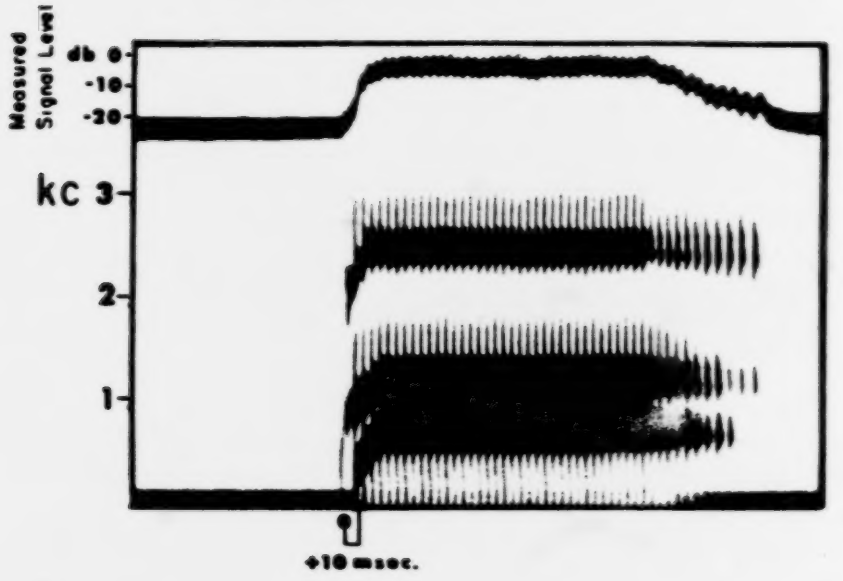
As part of our cross-language study of stop voicing, three series of synthetic speech patterns of the type illustrated by Figure 2 were generated, in which VOT was systematically varied over a 300 msec range, from a value of -150 (pulsing onset 150 msecs before the burst) to one of +150 (pulsing onset 150 msecs after the burst). These stimuli, in various appropriate random orders, were presented to speakers of three of the languages for which we

---

[3] Current work involving motion-picture photography of the glottis via fiber-optics indicates that this is regularly the case during production of voiceless aspirated stops, at least for English (Lisker et al., 1969).

Three Conditions of Voice Onset Time
Synthetic Labial Stops





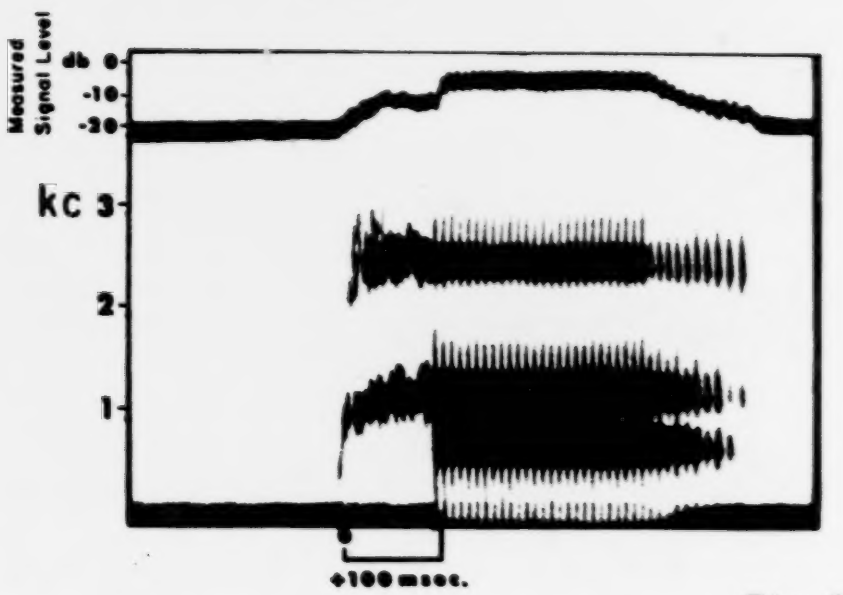


Fig. 2

had real-speech VOT measurement data. The languages chosen were English, Spanish, and Thai, the first two having two categories of stops each, and the third being representative of "three-category" languages.

Two kinds of data were gathered: labelling responses and something we called "discrimination" data. The procedure that yielded the labelling data involved asking subjects, native speakers of the languages mentioned, to name the initial stop of a stimulus by identifying it with one or another of the initial stops in their language. The labelling data obtained by presentation of our labial series of stimuli are shown in Figure 3. The responses that we took to reflect subjects' ability to discriminate between items of the stimulus set were collected by the following procedure. Stimulus triads were composed of two items that were identical in VOT value and a third differing from these by 20, 30, or 40 msecs along the same dimension. The order of presentation of members was random, with all possible orderings equally represented in the full set of triads submitted to the subjects. The subjects' task was to identify the "odd ball" as the first, second, or third member of the triad. Representative results are given in Figure 4, which shows discrimination functions for one English-speaking and one Thai-speaking subject.

In the procedure by which labelling responses were obtained as a function of VOT values, the possibility that subjects might recognize more categories than their language possessed was not taken into account. From the discrimination task and the data thereby obtained, it appeared that discriminability is not significantly better than chance for stimulus pairs which are categorially identical, but it increases sharply for pairs located near the boundary between categories along the VOT dimension. Insofar as the locations of these discrimination peaks differ for the speakers of different languages, and indeed sometimes for individual subjects, matching thereby the boundaries between linguistic categories established by the labelling tests, it would seem difficult to decide whether the failure to make subcategorial discriminations means that subjects cannot discriminate, or simply that they did not, given a test in which some comparisons were across a category boundary and others were not. One major purpose of some additional tests carried at the speech research laboratory of the Pavlov Institute near Leningrad was to learn whether speakers of Russian, a two-category language with voiced and voiceless unaspirated stops, can readily distinguish between stimuli which, for speakers of English, are categorially different but which for Russians are of the same category. There was, unfortunately, no opportunity to make VOT measurements of spoken Russian stops comparable in quantity with the large body of cross-language data presented in Lisker and Abramson (1964), but a modest amount of labelling and discrimination data was collected.

In order to obtain VOT labelling data from Russian speakers, the same synthetic speech stimuli tested previously with American, Puerto Rican, and Thai subjects were presented to a group of fifteen members of the Pavlov Institute staff. Their responses to these stimuli are given in Figure 5. Although Russian is a two-category language, with stops resembling those of Spanish and Hungarian, stimuli with VOT values greater than +80 were judged to begin with the cluster nx, i.e., the voiceless bilabial stop followed by a voiceless velar fricative. The labelling behavior of our Russian-speaking subjects may be compared with the data obtained from speakers of the

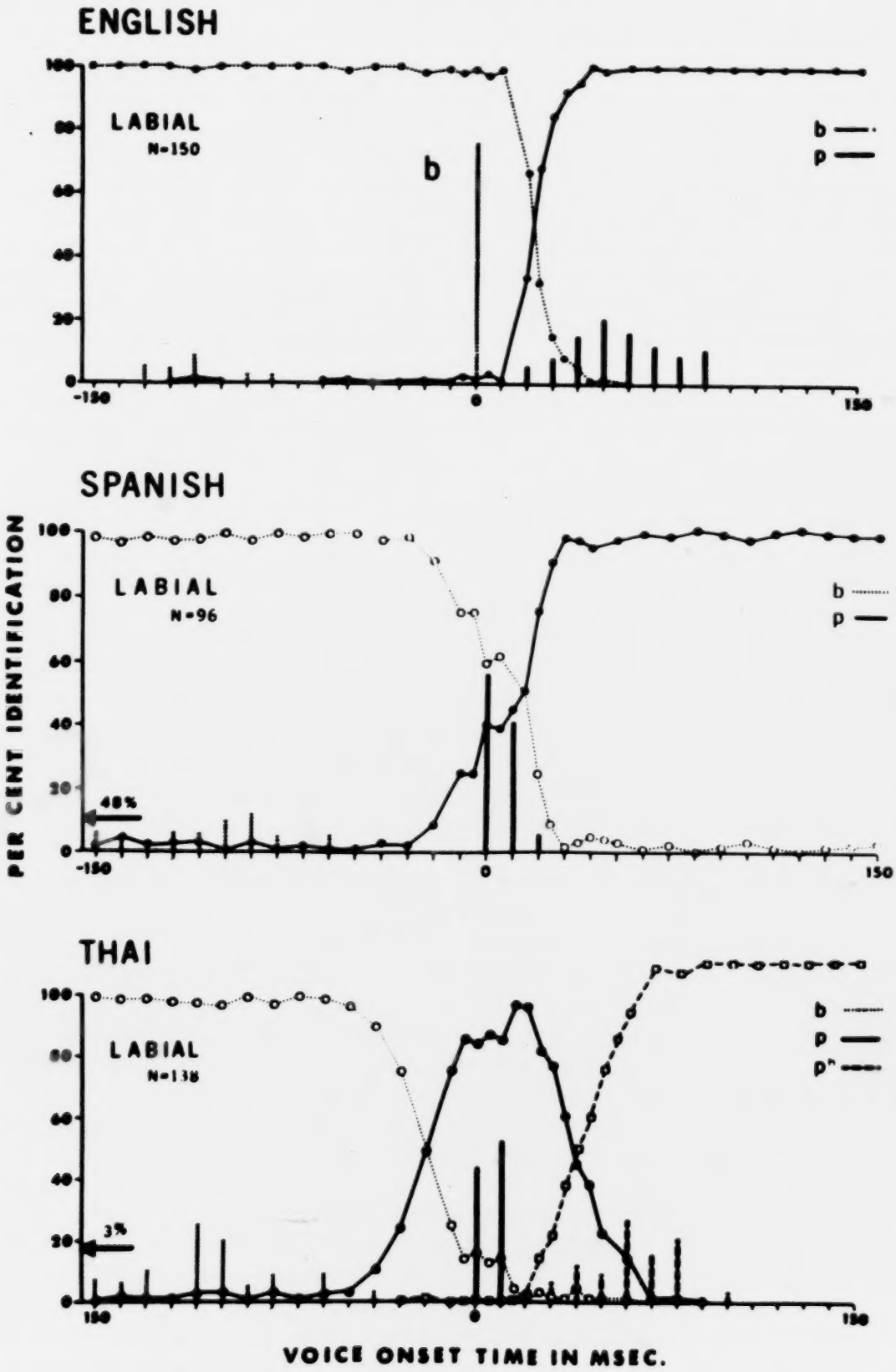

ENGLISH
LABIAL
N=150
b
b
p
100
80
60
40
20
0
-150
0
150
SPANISH
PER CENT IDENTIFICATION
LABIAL
N=96
b
p
48%
100
80
60
40
20
0
-150
0
150
THAI
LABIAL
N=138
b
p
pʰ
3%
100
80
60
40
20
0
150
0
150
VOICE ONSET TIME IN MSEC.

Fig. 3

LABIAL DISCRIMINATION

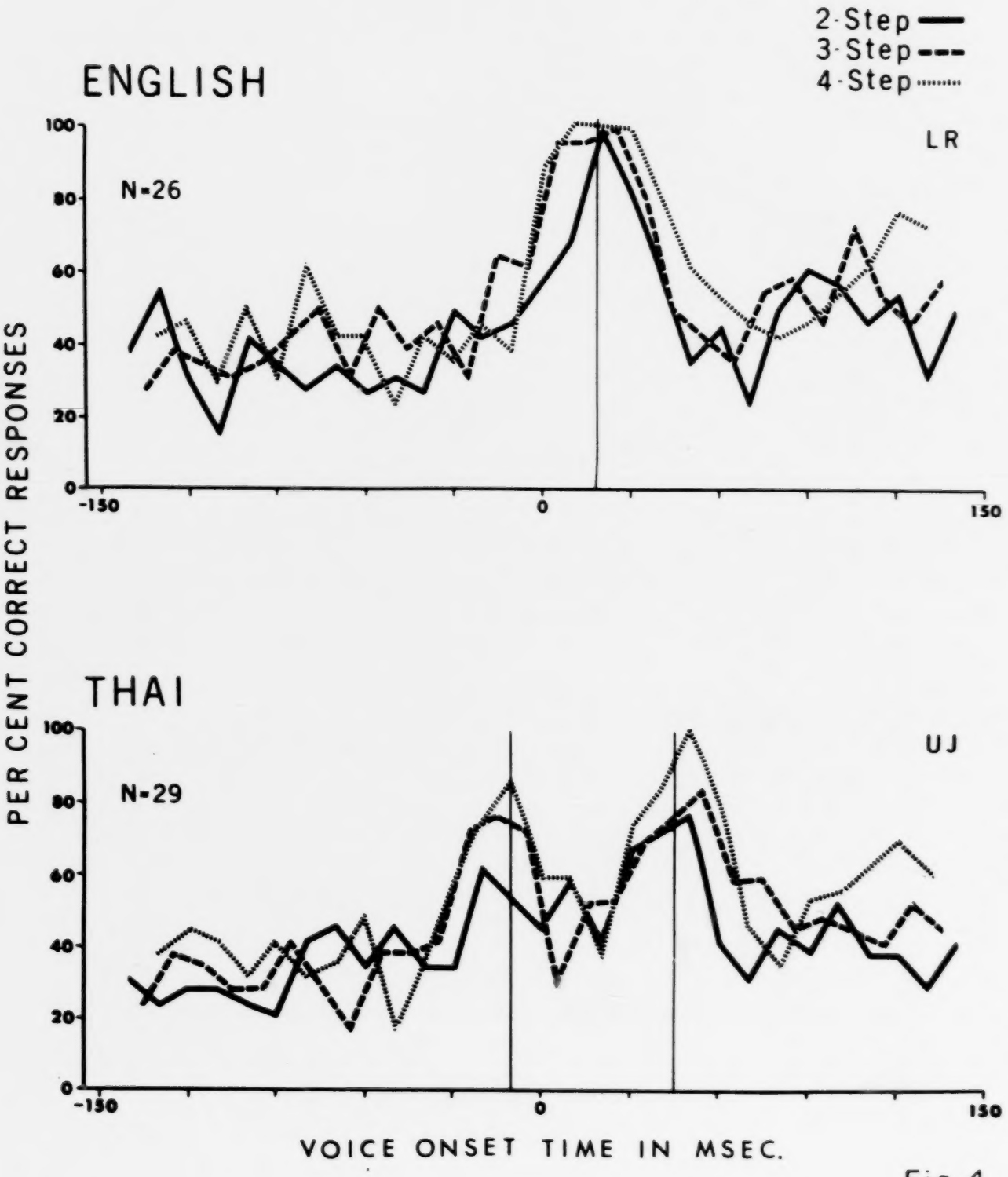


Fig. 4

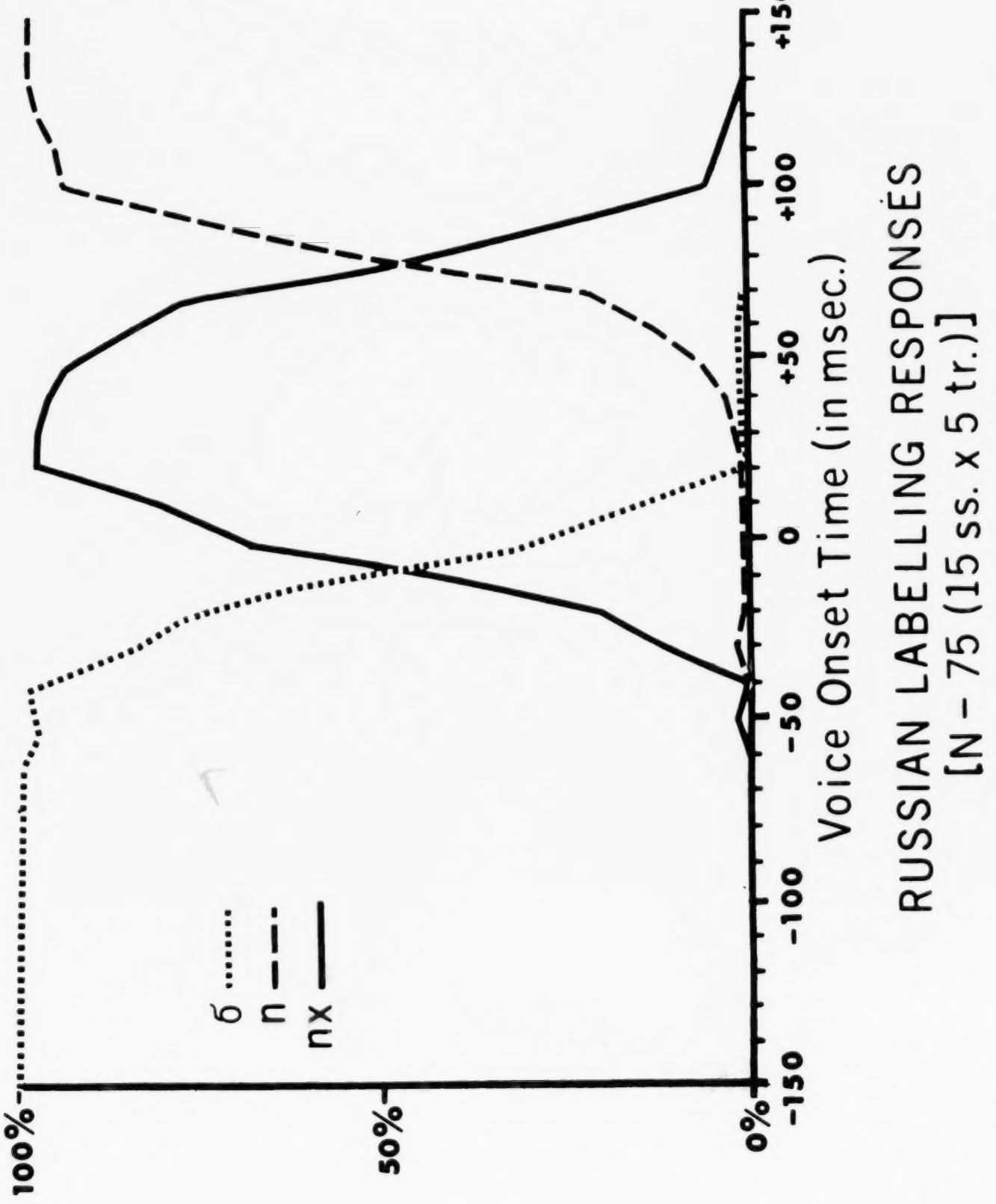

RUSSIAN LABELLING RESPONSES
[N – 75 (15 ss. x 5 tr.)]
Voice Onset Time (in msec.)
+150
+100
+50
0
-50
-100
-150
100%
50%
0%
δ
n
nx

Fig. 5

three other languages previously mentioned (Fig. 3). The differences from language to language are considerable, even if we allow for variations, probably minor, due to the fact that listening conditions could not be rigorously controlled. Presumably the fact that English speakers divide the VOT space into b and p categories at +25, while for Russians the crossover point dividing б from n is close to -10, is of significance, particularly in view of the fact that this difference is observed in actual speech production as well. At the same time it should be remarked that the match with production data is not always very good, specifically in the Spanish case and in the p-p^h boundary of Thai.

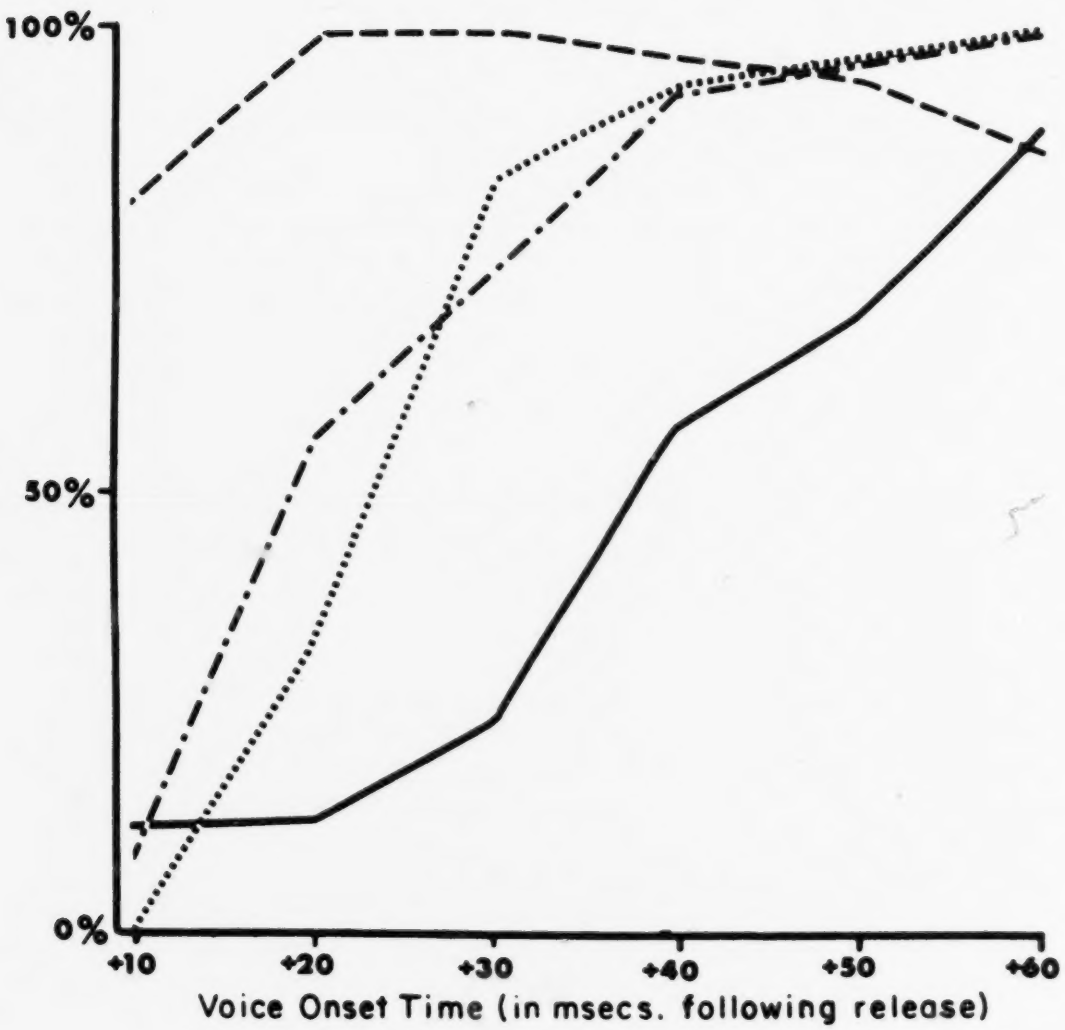
Since English speakers identify stimuli with VOT values in the range -10...+25 with items of lower (i.e., more negative) values, while Russians identify them with items having higher values, the question arises as to whether this difference means that acoustic cues available to both groups are assessed according to different strategies, or whether the cues to which one group attends are simply not available to the other. Is it the case, for example, that Russian listeners are quite capable of distinguishing between items at +20 and +50, although both are n for them, and that Americans can hear the difference between VOT values of -30 and 0, although both are labelled b? If one group is able to discriminate between stimuli which are categorially different only for the other, then the implication would be that there is a psychoacoustic basis for the category boundary.

In order to learn whether the boundary at VOT = +25 between English b and p is susceptible of detection by speakers of Russian, the following experiment was carried out. Items having VOT values of +10 and +60, both of them n, were presented to a group of Russian speakers who were instructed to assign different labels to them, i.e., to move a toggle switch one way for +10 stimuli and the other way for +60 stimuli. Subjects' success in learning this task was ascertained by presenting the two stimuli many times in random order, simultaneously registering their responses by means of an electromechanical recording system. It appeared that the test group was able to do significantly better than chance, with a majority of the six subjects tested getting above 90 percent correct in identifying the +10 items. Identification of the +60 stimuli was less good, but still better than 75 percent correct for all but one subject. Thus it seemed that the subjects as a group could both distinguish the two test stimuli and also apply two different labels to them in a reasonably consistent way.

A second labelling test was next constructed in which was presented a set of stimuli covering the range from +10 to +60 in steps of 10 msecs. The test subjects' task was to identify each stimulus by judging whether it was more similar to the +10 or the +60 item. Each stimulus was represented five times in the random order presentation, and the entire set of thirty items was administered to the same group of subjects repeatedly over several days.

The responses of our subjects in the labelling test just described are represented by the heavy solid line in Figure 6, which shows the percentage of +60 identifications as a function of VOT value. This labelling function is to be compared with the two functions representing the behavior of a group of English-speaking subjects tested subsequently in the United States:

Comparison of Russian and American Labelling of
Full and Restricted Ranges Along
the VOT Dimension



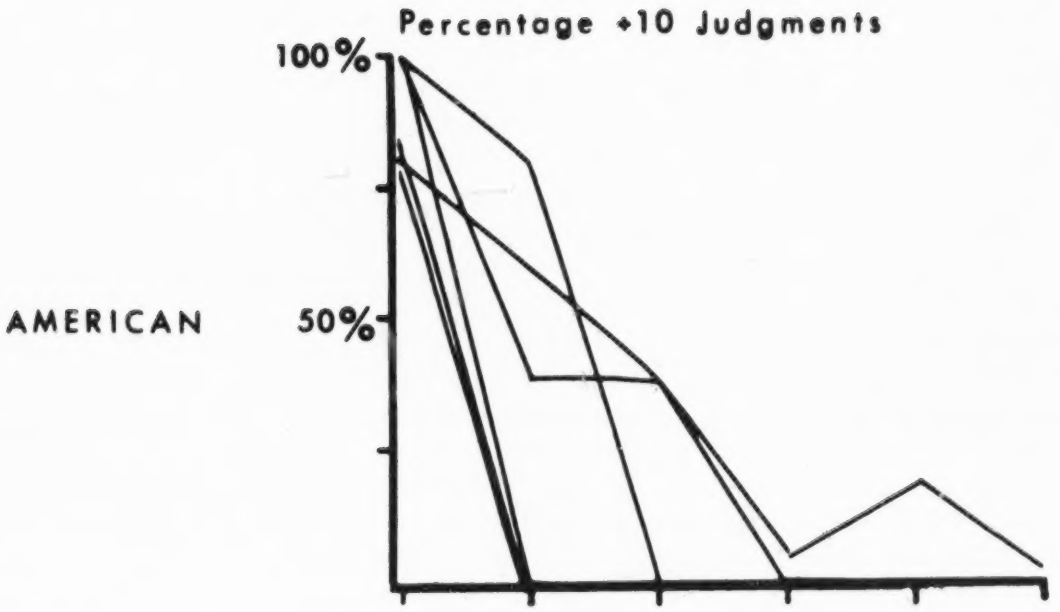
Key ——— : Russian identifications of restricted range (+10..+60) as +60

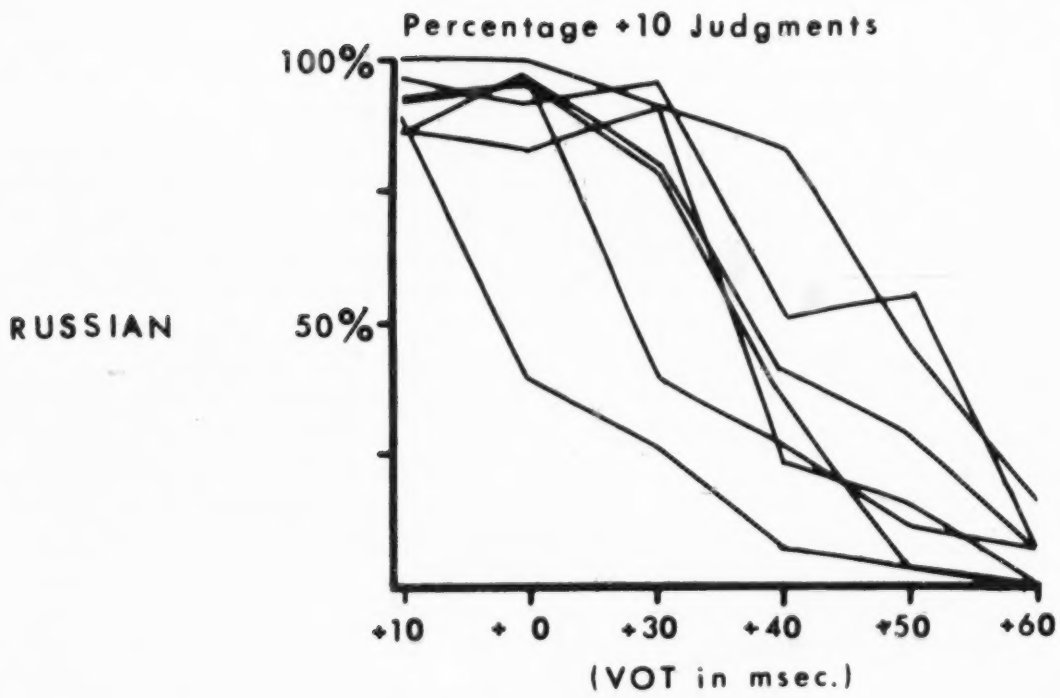
- - - - : Russian identifications of full range (-150..+150) as /n/

—·—·— : American identifications of restricted range (+10..+60) as +60

..... : American identifications of full range (-150..+150) as /p/

Fig. 6





Individual Variation in Identification of Stimuli with +10 as against +60

Fig. 7

the thin solid line in Figure 6 gives percentage p judgments derived from tests in which stimuli covering the full -150...+150 range were presented, while the thin dashed line in the same figure gives responses to the restricted set ranging from +10 to +60 along the VOT dimension. (The heavy dashed line, representing the Russians' n labellings for the full VOT range, is included in Figure 6 for ease of comparison.) It is apparent that the Russian and the American English data do not closely match, and one is tempted to believe that, by and large, the Russian listeners were making continuous, rather than categorical, judgments, i.e., that they were estimating the magnitude of the difference between a given stimulus and each of the standard stimuli rather than deciding whether or not it shared some feature with one of the standards. The Americans' judgments were very much the same in the two labelling tests; whether they were dividing the full VOT range into b and p categories or the restricted range by matching with the +10 as against the +60 stimuli, their judgments were evenly divided at about +20. The Russian judgments, by contrast, show a crossover somewhere between +30 and +40, i.e., at about the midpoint of the +10...+60 range; within this same range, of course, the curve representing the partition of the full VOT range into d and n categories does not approach the 50 percent value on the ordinate. When we look at the behavior of individual subjects, moreover, we find a marked difference in the degree of variability for the two groups; the American subjects are noticeably more alike in their division of the restricted range of stimuli than are the Russians (Fig. 7), who place their boundaries anywhere between +20 and +50. It may possibly be true that the single Russian listener who observed a crossover value near +20 was following the Americans' strategy, but the Russians as a group were certainly not attending to the same cues as the latter. On the other hand, it is possibly only accidental that the Russian crossover near +40 is very nearly at the midpoint of the +10...+60 range, so that we cannot be certain that they were estimating difference magnitudes rather than responding categorically to some acoustic cue. There is the possibility, moreover, that this crossover value near +40 is to be related to one of the crossover values determined for our Thai subjects in the full-range labelling test (Fig. 3). What we can be reasonably certain of, on the basis of our present data, is that the Russian listeners did not generally observe the boundary which served the American subjects in both labelling tasks. It would be appropriate to determine Russian crossover values for several additional VOT ranges, e.g., +20...+70 and +30...+80, that fall within the n category, in order to learn whether the crossover values remain fixed or tend to move with the range boundaries. In the first event, we should be in a position to assert that the Russian listeners were evaluating the stimuli, in categorical fashion, according to acoustic criteria other than those motivating the American listeners; in the latter event, we should have to suppose that a continuous kind of perception and comparison was being practiced.

REFERENCES

Abramson, A.S. and Lisker, L. 1970. Discriminability along the voicing continuum: Cross-language tests. Proceedings of the 6th International Congress of Phonetic Sciences, Prague, 1967, 569-573.

Dudley, H. 1940. The carrier nature of speech. Bell System Technical Journal 19.495-515.

Fry, D.B., Abramson, A.S., Eimas, P., and Liberman, A.M. 1962. The identification and discrimination of synthetic vowels. Language and Speech 5.171-189.

Liberman, A.M., Delattre, P.C. and Cooper, F.S. 1958. Some cues for the distinction between voiced and voiceless stops in initial position. Language and Speech 1.153-167.

Lisker, L. and Abramson, A.S. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. Word 20.384-422.

Lisker, L. and Abramson, A.S. 1967. Some effects of context on voice onset time in English stops. Language and Speech 10.1-28.

Lisker, L., Abramson, A.S., Cooper, F.S. and Schvey, M.H. 1969. Transillumination of the larynx in running speech. Journal of the Acoustical Society of America 45.1544-1546.

Lisker, L. and Abramson, A.S. 1970. The voicing dimension: Some experiments in comparative phonetics. Proceedings of the 6th International Congress of Phonetic Sciences, Prague, 1967, 563-567.

Mattingly, I.G. 1968. Experimental methods for speech synthesis by rule. IEEE Transactions, Audio. 16.198-202.

Mattingly, I.G., Syrdal, A.K., Liberman, A.M. and Halwes, T. 1969. Discrimination of F2 transitions in speech context and in isolation. Journal of the Acoustical Society of America 45.314-315.

Stevens, K.N., Liberman, A.M., Studdert-Kennedy, M. and Öhman, S.E.G. 1969. Crosslanguage study of vowel perception. Language and Speech 12, Part 1, 1-23.

Letter Confusions and Reversals of Sequence in the Beginning Reader: Implications for Orton's Theory of Developmental Dyslexia*

Isabelle Y. Liberman,[+] Donald Shankweiler,[++] Charles Orlando,[+] Katherine S. Harris,[+++] and Fredericka B. Berti[+++]

Reversals of letter order and orientation in reading and writing are generally thought to be of special importance for understanding developmental dyslexia. Interest in reversal errors stems largely from the work of S.T. Orton (1937) who viewed childhood dyslexia as one element of a developmental syndrome which has as its basis an anomaly of cerebral dominance. In forming this neurological conception of reading disability, Orton wished to establish a causal link between two observations: first, that children with reading disability tend to have poorly established or unstable lateral preferences, and secondly, that they tend to reverse letters and words in reading and writing. These difficulties were seen as related manifestations of a failure of one cerebral hemisphere to become dominant. This conception has been challenged by some workers in the field (Schonell, 1948; Burt, 1950; Vernon, 1960) and supported by others (Zangwill, 1960; Critchley, 1964). Though an extensive literature has been developed in the area, the question of a possible relation among reading reversals, motor ambilaterality, and cerebral dominance remains open.

In our view, the question is premature. The significance of reversals in dyslexia is unknown because the reversal phenomenon itself has not been studied systematically and a number of preliminary questions have not been fully answered. In the first place, it is not known how frequently and consistently reversals occur in beginning readers generally. Secondly, do reversals comprise a constant proportion of all errors? If so, it would be highly misleading to count the reversals a child makes without examining the other errors as well. Third, are the two types of reversals related? Orton (1937) had differentiated between "kinetic," or sequential, reversal of letter order and "static," or orientational, reversal of letter form. He did not doubt that they were closely associated, despite his own observation that they "vary markedly in their relative frequence and in the resistance which they offer to eradication by training" (p. 150). The problem of whether these are related phenomena raises a further question: are reversals solely a consequence of optical properties of the letter shapes? This could be true of reversible letters such as b and d, but another explanation is required to handle reversals of sequences.

---

*Paper to appear in Cortex.

[+]University of Connecticut, Storrs.
[++]Haskins Laboratories, New Haven, and University of Connecticut, Storrs.
[+++]Haskins Laboratories, New Haven, and City University of New York, Graduate Center.

Acknowledgment: This work was supported in part by a grant to the University of Connecticut from the U.S. Office of Education (co-principal investigators, I.Y. Liberman and D. Shankweiler).

Some investigators have viewed reversals in reading as a problem in visual perception, supposing that they are but one indicator of a more general perceptual immaturity, as manifested in such disabilities as poor form and space perception or defective memory for designs. A number of studies (Fildes, 1921; Gates, 1922; Kendall, 1948; Goins, 1958; Malmquist, 1958) have sought relationships between reading proficiency and various aspects of visual perception. The findings of these studies (critically reviewed by Benton, 1962) have been, for the most part, negative or equivocal. The work of Gibson and her colleagues (1962) has systematically explored the development of the discriminability of letters as optical shapes, by the use of letter-like forms which incorporate the basic features of letters of the alphabet. These researchers have assessed the relative difficulty of various transformations (including reversals) of the basic shapes. Valuable as this developmental study has been in clarifying the visual conditions of letter recognition, it has not dealt with the linguistic function of the letter shapes and, therefore, has limited relevance to our present problem of understanding errors in reading.

In view of these limitations of earlier work, we saw a need for an experimental study of reversal errors which would take into account the linguistic context as well as the optical properties of the stimuli and would investigate reversals in relation to the other errors the child makes when confronted with the printed word.

A number of researchers contemporary with Orton (see, for example, Monroe, 1932; Teegarden, 1933; Gates, 1933; Hildreth, 1934; Davidson, 1935; Hill, 1936; Wolfe, 1939; Bennett, 1942) concerned themselves either directly or tangentially with the nature of reversal errors in reading, but, for various reasons, their results are difficult to assess. Some considered only errors of orientation. Several discussed both types but did not treat the two separately in presenting their results. When they did consider them separately, they did not investigate further the relationship of the two kinds of error to each other or their relationships to other consonant and vowel errors occurring concomitantly. Special tests to measure reversal tendency have rarely been devised; most investigators culled the reversals from the children's performance on diagnostic reading paragraphs or word lists. Even when special tests were used, no attempt was made to assess the reliability of the findings or to adjust the observations for the opportunities available in the material for making various types of errors. Some studies took into account the effect of whole-word vs. single-letter presentation; usually, the possibility of different error frequencies in meaningful and nonsense material was not considered.

The same shortcomings listed above are found in more recent explorations of reversal error patterns in reading (Hermann, 1959; Tordrup, 1966). Thus, the relationship of sequence and orientation reversals to each other and to different aspects of reading mastery remains uncertain, as does the nature of the general error pattern in the disabled reader. This investigation was designed to provide a more systematic approach to these questions.

METHOD

Subjects

The subjects for this study were selected from the second grade of an elementary school system located in a small northeastern Connecticut town. A sixty-item word list (described below) was administered to the entire second grade population of the school system (N=59). Five children were eliminated as possible subjects. These included two with speech impairment, two who moved from the district before testing could be completed, and one who transferred to the school system after the initial segment of testing had begun.

The eighteen children chosen for further study comprised the full lower third of the remaining group in reading proficiency as determined by their total error score on the word list. School records indicated that none of the children had impaired hearing or uncorrected errors of refraction. Fifteen were boys and three were girls. Their ages ranged from 7.25 to 9.25 years (mean = 8.25 years). All tested within the normal range of intelligence according to the Wechsler Intelligence Scale for Children. (IQ range: 85 - 126; mean = 98.6).

Procedure

The following tasks were given to all the subjects in the same order on successive days:

1. Word List. List of sixty real-word monosyllables including a group of primer-level sight words, a group of non-sight words, and word-forming reversals of both types of words, where such were possible. The word list is shown as Table I.

Each word was printed in manuscript form with a black felt-tip pen on a separate 3" x 5" white index card. The cards were presented individually in the order in which they appear in the list in Table I, with the following instructions:

> I want you to read some words aloud for me. Some of the words are easy, and some are hard. If you don't know the word, try to sound it out. Do the best you can.

Responses were recorded on tape, as well as being transcribed during the administration, to check on the accuracy of transcription. Each child's responses were analyzed for reversals of sequence and of orientation, for consonant and vowel errors, and for total errors.

The word list was administered twice to each subject--once at the end of the school year and again in the first week of the following school year. Data from the two presentations were combined in scoring the responses of each subject but were available separately for assessment of test-retest reliability.[1]

---

[1] A list consisting of sixty CVC nonsense monosyllables was also administered to this group. Full discussion of this task will be reserved for a later paper, but certain of the data will be included here, where pertinent.

Table I

Word List

1.	of	21.	two	41.	bat
2.	boy	22.	war	42.	tug
3.	now	23.	bed	43.	form
4.	tap	24.	felt	44.	left
5.	dog	25.	big	45.	bay
6.	lap	26.	not	46.	how
7.	tub	27.	yam	47.	dip
8.	day	28.	peg	48.	no
9.	for	29.	was	49.	pig
10.	bad	30.	tab	50.	cap
11.	out	31.	won	51.	god
12.	pat	32.	pot	52.	top
13.	ten	33.	net	53.	pal
14.	gut	34.	pin	54.	may
15.	cab	35.	from	55.	bet
16.	pit	36.	ton	56.	raw
17.	saw	37.	but	57.	pay
18.	get	38.	who	58.	tar
19.	rat	39.	nip	59.	dab
20.	dig	40.	on	60.	tip

2. Gray Oral Reading Test, Form A. Administered by the standard procedure. Raw paragraph scores based on Gray's system of weighing time and number of errors were used to evaluate the subjects' performance, rather than grade-level equivalents.

3. Single-Letter Presentation (Tach.) List of 100 items in which a given letter was to be matched to one of a group of five, including four reversible letters in manuscript form (b, d, p, g) and one nonreversible letter (e) which was added as a reliability check. There were 20 such items for each letter. The order of the resultant 100 items was randomized, as was the order of the multiple-choice sequence for each item on the answer sheet. The standard was presented tachistoscopically for matching with one of the multiple-choice items on the answer sheets. Tachistoscopic exposure of the 2" x 2" slides of each letter was projected for 1/125 sec. in the center of a 9" x 12" screen mounted six feet in front of the subject. A brief training session was provided for each child.

Administration was the same as for the word list except for the instructions, which were as follows: "Here are some make-believe words. They are not real words; they are only pretend words. Read them aloud as well as you can."

Error Analysis of Word Transcription

The responses to the stimulus words were scored twice--first, from the transcription made at the time of the test administration, and second, from a separate transcription by another experimenter from the tape recording. Disagreements between scores were infrequent and were easily settled by invocation of the rules listed below.

1. Reversal of Sequence (RS). Scored when a word or a part of a word was read from right to left (e.g., when lap was read as [p ae l] or [pl eI]; form as [fram]).

2. Reversal of Orientation (RO). Scored when b, d, p, and g were confused with each other, as when bad was read as [d ae d], [p ae d], or [b ae g]. If bad was given as [d ae b], it was scored as a sequence error instead. Both types of reversal were scored when nip was read as [bIn].

3. Other Consonant Error (OC). Included all consonant omissions and additions as well as all consonant substitutions other than reversal of orientation. A response could contain both a sequence reversal and a consonant error, as in the case of the response [tr ae p] for the stimulus word pat. It could also contain both an orientation reversal and a consonant error, as in the case of the response [tr ae p] for the stimulus word tab. However, confusions among b, d, p, and g were scored only as reversals of orientation, not as consonant errors.

4. Vowel Error (V). Included all vowel substitutions, such as [pIg] for peg. A vowel error was not charged when a consonant error in the response forced a change in the pronunciation of the vowel, provided the vowel sound produced in the response was a legitimate pronunciation of the original printed vowel (response [r ae t] for the stimulus word raw).

5. Total Error. Simply the sum of all the preceding error types.

In general, the first concern was to assure that scoring was not falsely prejudiced in the direction of any given error category. To this end, certain additional rules were consistently invoked in the few instances when scoring was not immediately self-evident. The stimulus word was viewed in relation to its component printed vowels and consonants as written. The response word was considered phonetically, not in terms of the orthographic transcription of a possible target word. An exception was made when both the stimulus and target words contained vowel digraphs. As noted above, no vowel error was charged when a consonant error in the response produced a change in the pronunciation of the vowel, provided the original printed vowel would be sounded legitimately as read in its new consonant environment..

Several examples may serve to clarify the scoring. If tar was read as [tra], it was, of course, scored simply as a sequence reversal. If it was read as [treI], the response was scored as both a sequence reversal and a vowel error. Here no account was taken of the possible target word

tray. The response [tr ae p] would have been scored as a sequence reversal and a consonant error (for the addition of the p). In this case, no vowel error would be scored, since the original printed vowel would be sounded in this way in its new consonant environment (caused by the addition of the p).

Where the final consonant of the stimulus word was part of a vowel digraph as in the case of the word raw, substitutions for the w were viewed as consonant errors (e.g., raw read as [r ae t] or [r ae m]). Here, as was the case of [tr ae p] as a response for tar, no vowel error was charged, since both the stimulus word and the possible target word (ray) involved vowel digraphs ([ɔ] and [eI]).

RESULTS

Which Children Reverse?

The entire second-grade group was rank-ordered with respect to frequency of total errors on the word list. Nearly all of the reversal errors were found in children ranking in the lower third of the distribution. We, therefore, confined our study to these eighteen children who comprised the poorest readers.

It was apparent that, for most of the children, reversals accounted for only a rather small proportion of the total of misread letters. The means (as percentages of the total error) were 10 percent and 15 percent, respectively, for RS and RO, whereas other consonant (OC) errors accounted for 32 percent of the total, vowel (V) errors for 43 percent. Even among this group of poor readers, individual differences were fairly large: rates of RS ranged from 4 to 19 percent; rates of RO ranged from 3 to 32 percent. Thus it is clear that among poor readers reversals do not merely form a constant proportion of all errors: only some poor readers reverse. Certainly it is important to explore the other differences among children who do and do not make reversals of sequence and orientation.

Test-Retest Reliability

Since our method of reading assessment was untried, we were concerned to demonstrate its reliability. The test-retest reliability coefficient for the total error was .83; for OC errors, .69; for V errors, .64. Thus the general error rate among the children is stable, although they tend to give some redistribution of the errors among consonants and vowels. Both types of reversal errors give lower reliability coefficients (r_{12}=.43 for RS; r_{12}=.50 for RO), indicating that they are not highly stable error categories.

The Word List and Reading Fluency

Having presented our second-grade readers with a highly artificial task of reading monosyllabic words in isolation, we wished to know how performance on such a task related to a conventional measure of reading proficiency. For that purpose we selected the Gray's Oral Reading Test as the most appropriate test available. The obtained Pearson product-moment correlation coefficient (r) was .77 between total errors on our word list and score on Gray's paragraphs, demonstrating a high relationship

between error rates on isolated words and on connected text. This finding suggests that the problems of the beginning reader have more to do with the organization of syllables than with the scanning of larger chunks of text. If the subjects had done well on the word list, but poorly on the paragraphs, difficulty in scanning a line of text might have been implicated. Since performance on the connected text was so highly correlated with that on isolated words, the major source of difficulty for these children must be in decoding the words. Of course, decoding may not be the most important problem for poor readers at later stages of reading development.

Intercorrelations among the various measures are displayed in Table II.

Table II

Intercorrelation Matrix

	Reversed Sequence Errors	Reversed Orientation Errors	Other Consonant Errors	Vowel Errors	Single Letter(Tach) Errors	Gray's Oral	WISC IQ
Total Error	**73	28	**93	**91	19	**77	*56
RS Errors		03	**72	*56	14	45	34
RO Errors			09	20	04	15	17
OC Errors				**73	28	**71	*46
V Errors					08	**75	**59
Tach. Errors						01	16
Gray's Oral							38

Note: The table contains Pearson product-moment correlation coefficients. Decimal prints are deleted.

*p < .05
**p < .01

As a further indication of the stability of the major error categories computed from the word list, it is noteworthy that the OC error category correlated .73 with

V errors, and each correlated well with the independent measure of reading proficiency, the Gray paragraphs (OC errors x Gray paragraphs, r = .71; V errors x Gray paragraphs, r = .75).

Lack of Correlation Between the Two Types of Reversals

Although Orton (1937) distinguished between reversals of letter sequence and letter orientation, he and his successors tended to assume that both are manifestations of the same underlying disturbance, namely, a failure to develop a consistent automatic left-to-right pattern of scan. Having considered the two types of reversal separately, we find no support whatever for supposing that they have a common cause: RS and RO were wholly uncorrelated (r = .03).[2] That means, of course, that an individual's frequency of misordering letter sequences is entirely unpredictable from his frequency of confusing geometrically ambiguous letters.

The two types of reversals, moreover, correlate quite differently with other measures. Inspection of the matrix of intercorrelations (Table II) reveals that RS is significantly correlated with OC and V, whereas neither of these is significantly correlated with RO. There is then clearly no justification for grouping the two types of reversal errors together.

Reversals in Relation to Other Errors

Table III gives frequencies of errors (for RS, RO, OC, and V) each percentaged according to the opportunities for errors of that type. This tabulation permits us to compare the relative frequencies of the various types of errors. First, we see, in agreement with classroom experience, that letters representing vowels are far more often misread than consonants.

Table III

Errors as a Function of Opportunity

	Reversed Sequence	Reversed Orientation	Other Consonant	Vowel	Single Letters (Tach.)
Errors	136	202	447	598	133
Opportunities	2160	1584	4284	2232	1800
Percent	6.3	12.7	10.4	26.8	7.4

[2]After preparation of this manuscript was completed, we discovered a study by Lyle (Lyle, J.O., Reading retardation and reversal tendency: A factorial study. Child Development, 1969. 40, 833-843) in which, in agreement with our findings, an absence of correlation between RS and RO was noted.

Reversals of orientation (RO) have a greater relative frequency of occurence than sequence reversals (RS) or other consonant errors (OC).[3] Thus, optically reversible letters appear to be a source of special difficulty for some poor beginning readers. It is important to note, however, that the problem with reversible letters is specific to reading; when the task is to identify these letters individually, even at rapid exposures, few errors occurred (Mean = 7.4%). Clearly then, the fact that these letters are a special source of difficulty in reading cannot be regarded simply as a problem in form perception.

Reversed Orientation of Letters: The Nature of the Confusions

It is of some interest to examine closely the particular confusions among letters which are formed by 180° transformations of the same basic shape. Confusions among the four reversible letters are presented as a matrix in Table IV. The matrix shows, with respect to each letter, the frequency with which it was replaced by another letter. Each row in the matrix refers to letters occurring in the word list and each column refers to the responses given by the children in oral reading. These frequencies are expressed as percentages of the total occurrences of each letter in the list (i.e., in terms of opportunities for error).

Table IV

Confusions Among Reversible Letters in Word List

Percentages Based on Opportunities

	Obtained					
Presented	b	d	p	g	Total Reversals	Other Errors
b	----	10.2	13.7	0.3	24.2	5.3
d	10.1	----	1.7	0.3	12.1	5.2
p	9.1	0.4	----	0.7	10.2	6.9
g	1.3	1.3	1.3	---	3.9	13.3

Confusion of b and d is the reversal most commonly mentioned in the literature and was interpreted by Orton (1937) as an instance of "sinistrad scan." It will be seen from Table IV, however, that in this group of children, p is given for b more frequently than is d. Indeed, in the table

[3] RO in Table III is lowered by the inclusion of g, which, as shown in Table IV, produces very few confusions with b, d, or p. When based only on these three truly reversible letters, RO increases to 15.5 percent.

as a whole, there were slightly fewer occurrences of 180° transformations in the horizontal plane (b to d, for example) than in the vertical plane (b to p, for example). This does not support the view that letter reversals are attributable to reversed direction of scan.

We also learn from the table that errors are essentially confined to confusions among b, d, and p. The letter g is, of course, a distinctive shape in all type styles, but it was included among the reversible letters because, historically, it has been treated as a reversible letter. It indeed becomes reversible when hand printed with a straight segment below the line. (Even in manuscript printing, as was used in preparing the stimulus materials for this study, the tail of the g is the only distinguishing characteristic.)

Concerning the confusions among b, d, and p, the truly reversible consonants, most errors involved a single 180° transformation about the vertical axis or the horizontal axis, but not both. Presumably, the presence within the alphabet of equivalent or near-equivalent optical shapes is one determinant of confusions among the letters b, d, and p, and by the same reasoning, the lack of congruence between these and g accounts for the rarity of the g substitution for b, d, or p. This conclusion is also supported by the relatively small frequencies of non-reversal errors (i.e., substitutions outside the set defined by the matrix) for b, d, and p in contrast to g.

Can we make sense of the pattern of the actual distribution of errors among the letters which differ in orientation but not in form? Table IV shows that at least twice as many errors occurred on b as on d or p. We may speculate on why this should be so. It may be relevant that b offers two opportunities to make a single 180° transformation whereas d and p offer only one. But there could also be a phonetic reason for the greater error rates on b, in that it offers the reader two opportunities to err by a single articulatory feature (place or voicing) whereas d and p offer only one opportunity to make a single feature error. This would be consistent with the finding that errors in perception of spoken consonants tend to differ from the presented consonant in only one feature (Miller and Nicely, 1955).

The present study gives no clear basis for choosing between these alternative interpretations.

We had also presented to the same children a list of pronounceable nonsense syllables with instructions that these were "pretend" words and that the children should attempt to sound them out as best they could. As expected, many more errors occurred on these than on real words, and the children tended to err by converting a nonsense syllable into a word.[4]

[4]When children read real words, their errors also are directed toward producing real words. Furthermore, an examination of the error distribution by individual words in the list shows that errors are concentrated on words where a well-known word is available as an error possibility. This tendency seems to affect the production of both RS and RO. However, since our stimulus list was not designed to investigate this factor specifically, we will postpone further speculation of the mechanisms involved for further research.

Again, g was rarely confused with the other three. However, the distribution of b errors was different from that which has been obtained with real words in that b - p confusions occurred only rarely. A check of the number of real words that can be made by reversing b in the two lists revealed no fewer opportunities to make words by substitution of p than by substitution of d, indeed, the reverse was the case. This result suggests that the nature of substitutions, even among reversible letters, is context dependent and therefore not an automatic consequence of the property of optical reversibility.

We may then ask whether confusions among b, d, and p occur outside of word context. When reversible letters were presented tachistoscopically as isolated shapes, relatively few misidentifications occurred (see Table III) and, moreover, RO and Tach. are uncorrelated (Table II). Thus, the characteristic of reversibility is not a sufficient condition for confusion.

We may conclude that, for whatever reason, b is significantly more often misread than other consonants. The fact that the errors b, d, and p tend to be confusions within the set suggests that the possibility of generating another letter by simple 180° transformation is a relevant factor in producing this high error rate. On the other hand, we have seen that RO's are, as are OC and V errors, context dependent and thus reflect the workings of linguistic processes as well as purely visual ones.

DISCUSSION

Reversals of letter sequence and letter orientation occurred in significant quantity only among the poorer readers in our groups of second graders. Even within the lower third of the class, they accounted for only 10 percent and 15 percent, respectively, of the total of misread letters, whereas other consonant errors accounted for 32 percent of the total and vowel errors accounted for 43 percent. Viewed in terms of opportunities for error, RO's occurred somewhat more frequently than other consonant errors, RS's definitely less frequently. Test-retest comparisons showed that, whereas other reading errors are rather stable, reversals, and particularly RO's, are unstable. Individual differences in reversal tendency were also large. Thus the indications from the analysis of variability, both intrasubject and intersubject, are that reversals do not form a constant proportion of all errors; only certain poor readers reverse, and it will be important to explore other differences between children who do and do not have reversal problems.

Although we have stressed that reversals of either type do not account for a large proportion of the total error in most of the children we have studied, it may be that reversals loom larger in importance in certain children with particularly severe and persistent reading difficulty. Our clinical experience suggests that this may be so, and we intend to explore the question in future research.

Examination of the intercorrelations among various reading errors showed that the two types of reversals are wholly uncorrelated. This is a finding of considerable interest since both were considered by Orton and subsequent investigators to be manifestations of an underlying tendency to reverse the direction of scan. That view cannot easily be reconciled with two additional findings: first, among reversible letters, vertical reversals occurred with as great frequency as horizontal reversals. Second,

confusions among reversible letters rarely occurred when these letters were presented singly, even when briefly exposed.

We investigated the relationship between reversals of both types and other errors in reading syllables. The findings are clear cut: individual error rates on vowels and consonants correlated highly with each other (another indication of the stability of our test), and each also correlated highly with the Gray's Oral Reading Test, an independent measure of reading proficiency. Frequency of RS correlated moderately with frequencies of other errors in reading the words and with the measure of reading proficiency, whereas RO frequency yielded weak and nonsignificant correlations with every other measure.

An analysis of the nature of substitutions among reversible letters (b, d, p, g) was carried out. This showed that the possibility of generating another letter by a simple 180° transformation is a relevant factor in producing a relatively high rate of confusion among these letters, in agreement with conclusions reached by Davidson (1935) and by Gibson and her associates (1962).

At the same time, other observations indicate the importance of linguistic determinants: differences in the pattern of confusions among b, d, p, and g in real words and nonsense words show that misperceptions even of reversible letters are context dependent and not merely an automatic consequence of optical reversibility. Moreover, the substitutions tended to differ from the presented consonant in one phonetic feature. Finally, relatively few confusions of these letters occurred when they were presented in isolation rather than in word context. All of these observations point to the conclusion that the characteristic of reversibility is not by itself a sufficient condition for confusion. (See Kolers, 1970, for a general discussion of how perception of letters and words differs from perception of nonlinguistic forms).

Further exploration of the linguistic determinants of children's reading errors is likely to be profitable. In this connection, the high correlation between reading proficiency on the word list and the paragraphs of the Gray's Oral Reading Test suggests that for the beginning reader, at least, an analytic test consisting of monosyllables can be substituted for a test employing connected text. We consider this an important finding because it indicates that a major part of the difficulties of the beginning reader has to do with the rules governing the synthesis of syllables from combinations of letter segments, rather than with strategies for scanning connected text.

This conclusion is supported by the results of a direct comparison of rate of scan in good and poor reading children (Katz and Wicklund, in press; see also Sternberg, 1967). It was found that both good and poor readers require 100 msec. longer to scan a three-word sentence than a two-word sentence, indicating that both have equivalent scanning rates and suggesting that they differ instead in some aspect of the decoding process.

SUMMARY

The pattern of errors of second-grade pupils in reading isolated words was analyzed, particularly with respect to reversals of letter sequence and

letter orientation. These occurred in significant quantity only among the poorer readers in the school class. The two types of reversals were uncorrelated and, therefore, cannot reflect a single process as Orton had implied. Sequence reversals were more closely related to other kinds of reading errors than were orientation reversals. The linguistic context as well as optical reversibility of letters is a determinant of confusions in letter orientation. Reading ability assessed by the analytic test composed of isolated words was highly correlated with reading proficiency on a conventional paragraphs test. This suggests that the problems of the beginning reader have more to do with word construction than with strategies for scanning connected text.

REFERENCES

Bennett, A. (1942) An analysis of errors in word recognition made by retarded readers. J. Educ. Psychol. 33. 25-38.

Benton, A.L. (1962) Dyslexia in relation to form perception and directional sense. In Reading Disability, ed. by J. Money. Johns Hopkins Press, Baltimore.

Burt, C. (1950) The Backward Child (3rd ed.). University of London Press.

Critchley, M. (1964) Developmental Dyslexia. Heineman, London.

Davidson, H. (1935) A study of the confusing letters B, D, P, and Q. J. Genet. Psychol. 47, 458-468.

Fildes, L.G. (1921) A psychological inquiry into the nature of the condition known as congenital word-blindness. Brain 44, 286-307.

Gates, A.I. (1922) The psychology of reading and spelling with special reference to disability. Teachers College Contributions to Education, No. 129.

Gates, A.I. (1933) Reversal Tendencies in Reading: Causes, Diagnosis, Prevention and Correction. Teachers College Bureau of Publications, Columbia University.

Gibson, E.J., Gibson, J.J., Pick, A.D., and Osser, R. (1962) A developmental study of the discrimination of letter-like forms. J. Comp. Physiol. Psychol. 55, 897-906.

Goins, J.P. (1958) Visual perceptual abilities and early reading progress. University of Chicago Supplementary Education Monographs, No. 87.

Hermann, K. (1959) Reading Disability: A Medical Study of Word Blindness and Related Handicaps. Charles C. Thomas, Springfield, Ill.

Hildreth, G. (1934) Reversals in reading and writing. J. Educ. Psychol. 25, 1-20.

Hill, M.B. (1936) A study of the process of word discrimination in individuals beginning to read. J. Educ. Res. 29, 487-500.

Katz, L. and Wicklund, D.A. (in press) Word scanning rate for good and poor readers. J. Educ. Psychol.

Kendall, B.S. (1948) A note on the relation of retardation in reading to performance on a Memory-for-Designs Test. J. Educ. Psychol. 39, 370-373.

Kolers, P.A. (1970) Three stages of reading. In Basic Studies on Reading, ed. by H. Levin and J. Williams. Harper and Row, New York.

Malmquist, E. (1958) Factors Related to Reading Disabilities in the First Grade of the Elementary School. Almqvist and Wiksell, Stockholm.

Miller, G. and Nicely, P.E. (1955) An analysis of perceptual confusions among some English consonants. J. Acoust. Soc. Amer. 27, 338-352.

Monroe, M. (1932) Children Who Cannot Read. University of Chicago Press.

Orton, S.T. (1937) Reading, Writing and Speech Problems in Children. W.W. Norton, New York.
Schonell, F.J. (1948) Backwardness in the Basic Subjects. (4th ed.) Oliver and Boyd, Edinburgh.
Sternberg, S. (1967) Two operations in character recognition: Some evidence from reaction-time measurements. Perception and Psychophysics 2, 45-53.
Teegarden, L. (1933) Tests for the tendency to reversal in reading. J. Educ. Res. 27, 81-97.
Tordrup, S.A. (1966) Reversals in reading and spelling. The Slow Learning Child 12, 173-183.
Vernon, M.D. (1960) Backwardness in Reading. (2nd ed.) Cambridge University Press.
Wolfe, L.S. (1939) An experimental study of reversals in reading. Amer. J. Psychol. 52, 533-561.
Zangwill, O.L. (1960) Cerebral Dominance and its Relation to Psychological Function. Oliver and Boyd, Edinburgh.

Perception of Dichotically Presented Steady-State Vowels as a Function of Interaural Delay*

Emily Kirstein+
Haskins Laboratories, New Haven

When stop-vowel syllables are presented dichotically with a temporal delay between syllables at the two ears, the lagging stop is recalled more accurately on the average then the one which leads (Studdert-Kennedy et al., 1970; Lowe et al., 1970). Porter et al. (1969) have reported finding a slight lead advantage rather than a lag advantage when the competing stimuli were steady-state vowels rather than stop consonants. They interpret the difference between stops and vowels as evidence that the lag effect in recall of dichotic stimuli is associated with special speech decoding processes. They attribute the absence of the effect for vowels to the vowels' being treated perceptually like nonspeech stimuli.

There is considerable evidence that steady-state vowels can be perceived either in a speech mode or nonspeech mode. For example, Spellacy and Blumstein (in press) claim that dichotically presented vowels will give either a left-ear advantage like nonspeech stimuli or a right-ear advantage like stop consonants depending on whether the test context induces subjects to listen for speech or nonspeech stimuli.

If the lag effect is associated with perception in the speech mode, then it should be possible to obtain a lag effect for isolated steady-state vowels under conditions which induce the subjects to perceive the vowels as speech. I have data which indicate that this is so. Figure 1 compares two groups of subjects. The ten naive subjects were people who had never taken a dichotic test. The ten experienced subjects had previously taken at least one dichotic test in which they identified vowels in stop-consonant vowel syllables. The stimuli in the present test were three synthetic steady-state vowels [e], [a], [ɔ] present in pairs at opposite ears with delays of 10, 30, 50, 70, or 90 msec between onsets of stimuli at the two ears. The subjects' task was to identify both vowels on each trial and to indicate which vowel sounded clearer by recording the clearer stimulus in the first column of the answer sheet. The graphs show the pattern of first responses for the two groups. Negative time values refer to lagging stimuli.

The naive subjects showed a slight lead advantage and a slight right-ear effect, neither effect significant. The experienced subjects showed both a right-ear advantage and a lag advantage. The two groups differ

---

*Paper presented at the eightieth meeting of the Acoustical Society of America, Houston, 3-6 November 1970.

+Also, University of Connecticut, Storrs.

ISOLATED VOWELS – FIRST RESPONSES

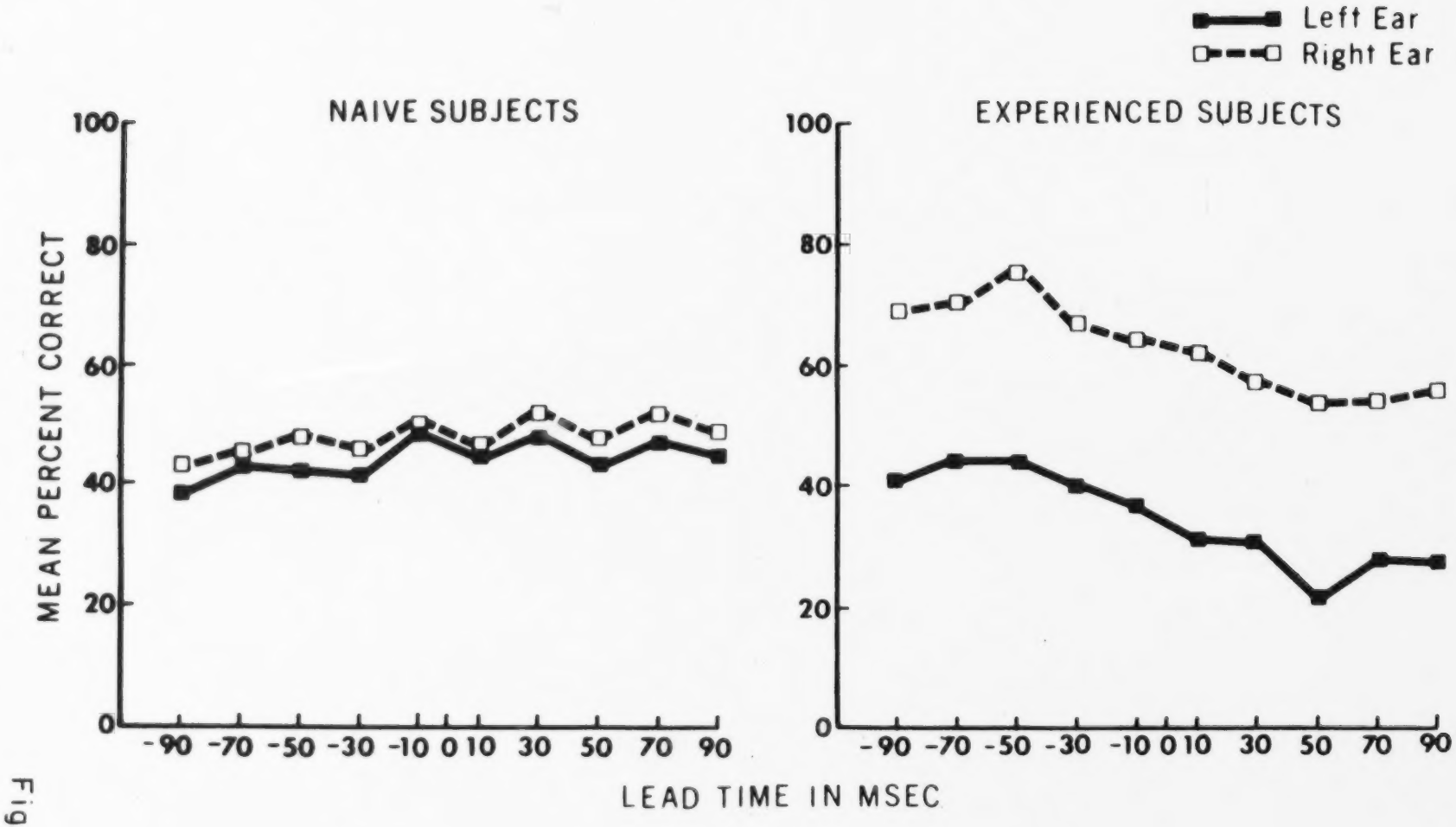


Fig. 1

ISOLATED VOWELS
LATERALITY EFFECT vs LAG EFFECT

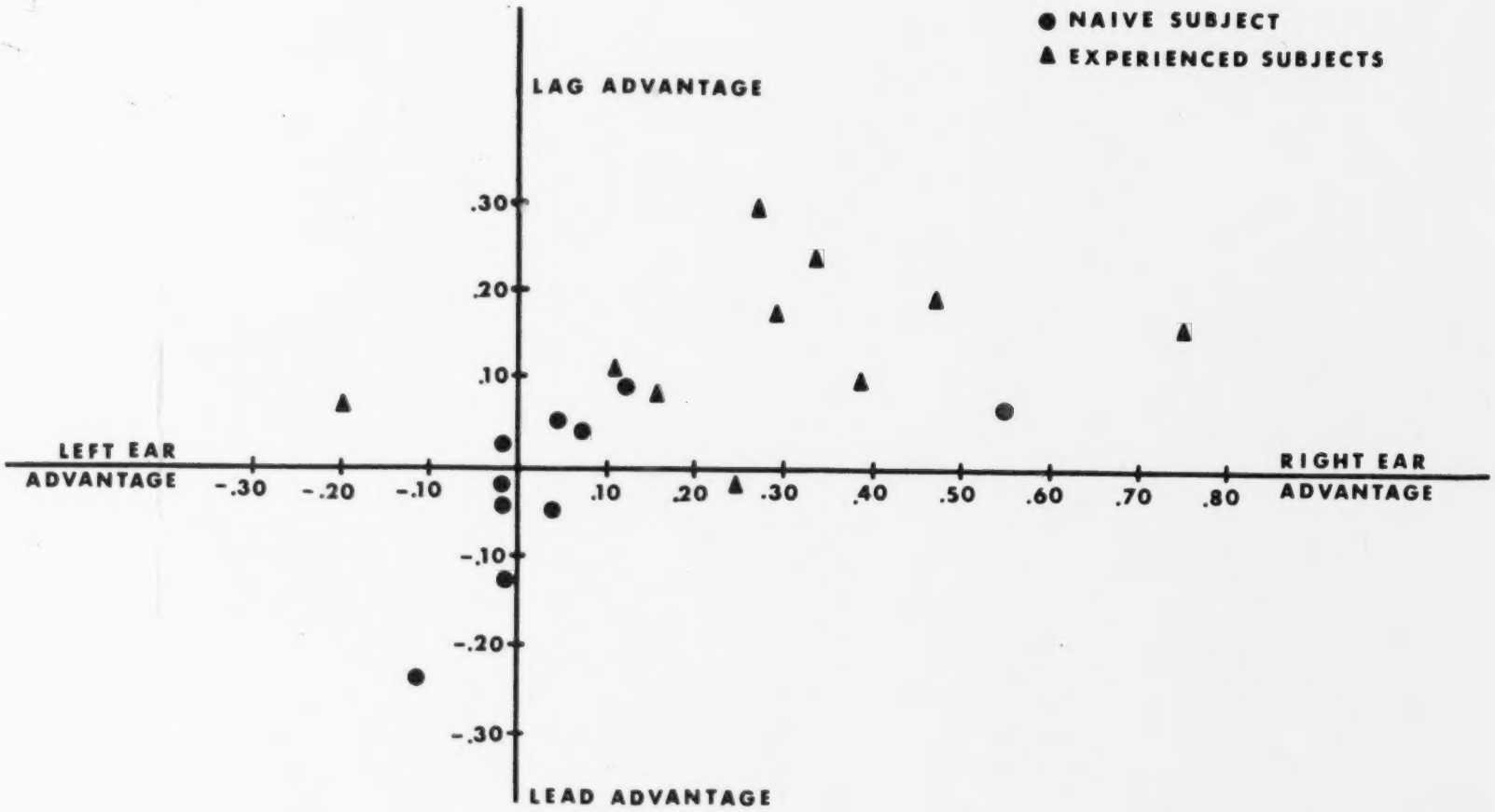


Fig. 2

significantly from each other. Apparently previous experience in dichotic tests involving vowels in CV syllables can shift listeners into a speech mode for isolated vowels, and this shift affects their time preference as well as their ear preference.

Indices of the magnitude of the lag effect and ear effect were computed for each subject based on first responses. A scatter plot of the ear effect vs. the lag effect is shown in Figure 2. Each naive subject is represented by a dot, each experienced subject by a triangle. An association between the right-ear effect and the lag effect was found not only when the two groups are compared with each other but also when individuals are compared within each group. The Spearman rank correlation coefficient between the right-ear effect and lag effect was +.71 for the ten naive subjects, which is significant with $p < .025$ on a one-tailed test. For the experienced subjects the correlation between the two effects is positive but not significant (.50).

For competing stop consonants most subjects show both a right-ear advantage and a lag advantage. I have not found any systematic relationship between the magnitude of the lag effect and right-ear effect for dichotically presented stops. The failure to find a positive correlation for stops suggests that the lag effect and right-ear effect in dichotic listening involve independent processes. Both effects may nevertheless depend upon the stimuli being perceived as speech. The results I have reported here for isolated vowels can be rationalized by the assumption that vowels can be analyzed by both speech and nonspeech processors. Both effects would occur when the speech mode predominates. Neither the lag effect nor the right-ear effect would be obtained when the nonspeech mode predominates. The significant correlation between the two effects for naive subjects suggests that individual subjects are shifting back and forth between the speech and nonspeech modes within the same test. The lack of lateralization typically reported for dichotic vowels may mean that the vowels are analyzed inconsistently by speech processors in the left hemisphere and nonspeech processors in the right hemisphere.

In summary, the finding of a positive correlation between the lag effect and right-ear effect for steady-state vowels supports the conclusion that the lag effect is a manifestation of processing in the speech mode.

REFERENCES

Lowe, S.S., J.K. Cullen, C. Thompson, C.I. Berlin, L.L. Kirkpatrick, and J.T. Ryan. 1970. Dichotic and monotic simultaneous and time-staggered speech. J. Acoust. Soc. Amer. 47, 76 (A).

Porter, R., D. Shankweiler, and A.M. Liberman. 1969. Differential effects of binaural time differences in perception of stop consonants and vowels. Proc. 77th Annual Convention of the Amer. Psychol. Assn.

Spellacy, F. and S. Blumstein. In press. The influence of language set on ear preference in phoneme recognition. Cortex.

Studdert-Kennedy, M., D. Shankweiler, and S. Schulman. 1970. Opposed effects of a delayed channel on perception of dichotically and monotically presented CV syllables. J. Acoust. Soc. Amer. 48, 599-602.

Perceptual Competition Between Speech and Nonspeech*

Ruth S. Day+ and James E. Cutting+
Haskins Laboratories, New Haven

Ear advantages are reported throughout the dichotic listening literature. When different messages are presented to each ear at the same time, the information presented to one ear is identified more accurately than the information presented to the other ear. In order to get an overview of the ear advantage results, consider the shorthand summary shown in Figure 1. When both inputs are speech (S/S), there is a right-ear advantage: stimuli presented to the right ear are identified more accurately than those presented to the left ear. This result has been obtained for a wide variety of speech stimuli, including digits (Kimura, 1961), words (Borokowski, Spreen, and Stutz, 1965), and consonant-vowel syllables (Shankweiler and Studdert-Kennedy, 1967). When both inputs are nonspeech (NS/NS), there is a left-ear advantage: stimuli presented to the left ear are identified more accurately. This result has been shown for various nonspeech stimuli, including melodies (Kimura, 1964), sonar signals (Chaney and Webster, 1966), and environmental noises (Curry, 1967).

Given that S/S yields a right-ear advantage and NS/NS yields a left-ear advantage, what happens when we present speech to one ear and nonspeech to the other ear at the same time? Will there be any ear advantage? One plausible prediction is that performance will be best when each stimulus is presented to its "proper" ear, that is, when nonspeech goes to the left ear and speech goes to the right ear (NS_L/S_R). This is a reasonable prediction if indeed the right-ear/left-hemisphere system processes speech and the left-ear/right-hemisphere system processes nonspeech, as suggested by Kimura (1967). The present study was designed to study the perception of speech and nonspeech pairs in the dichotic listening situation (S/NS).

EXPERIMENT I - IDENTIFICATION TASK

Method

Stimuli. The speech stimuli were the syllables /ba, da, ga/. They were real speech samples produced by one of the authors (RSD). The nonspeech stimuli were sine wave tones of 500, 700, and 1000 Hz. All stimuli were 300

---

*Paper presented at the eightieth meeting of the Acoustical Society of America, Houston, 3-6 November 1970.

+Also, Department of Psychology, Yale University, New Haven.

Summary of Ear Advantage Results from Previous Dichotic Listening Studies That Used Two Speech Stimuli (S/S) or Two Nonspeech Stimuli (NS/NS); Situation Studied in the Present Study (S/NS)

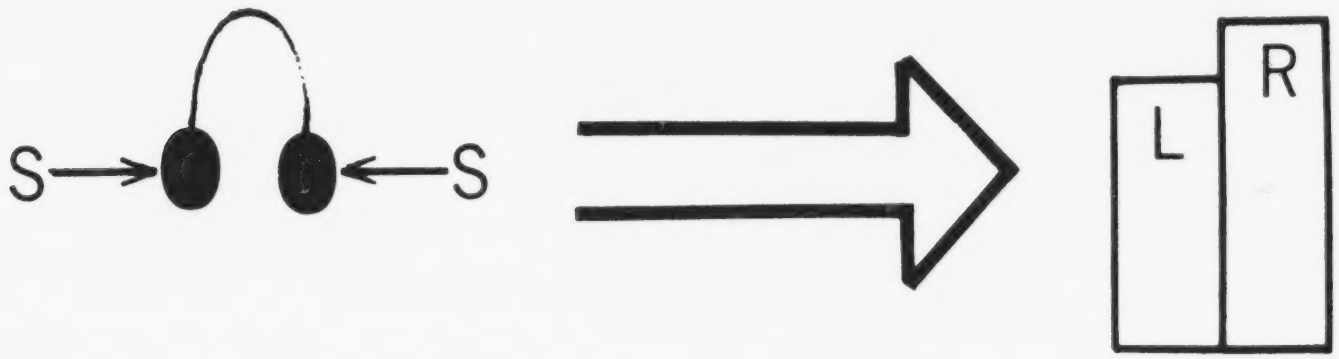


Fig. 1

msec in duration. The intensity envelopes of the tone stimuli were shaped to match those of the speech stimuli.

Paradigm. On each trial, one of the speech stimuli was paired with one of the tone stimuli. The assignment of stimulus type to ear varied randomly over trials, and all of the appropriate counterbalancing measures were taken into account. The onsets of the dichotic pairs were aligned using the pulse code modulation system at the Haskins Laboratories (Mattingly and Cooper, 1969). On some trials, both the speech and tone began at the same time. A sample pair with this 0-lead time is shown in Figure 2. On the remaining trials, one of the stimuli began first, by 25, 50, 75, 100, 150, or 200 msec. This paradigm enabled us to assess the role of temporal cues in perceiving the dichotic S/NS stimuli.

Subjects. Ten Yale University students served as subjects. All were right handed, were native English speakers, and had no history of hearing trouble. Each was tested individually.

Pretest. Before beginning the main experiment, each subject successfully completed a binaural identification test for the six stimuli.

Procedure. The subject was told that a speech stimulus would be presented to one ear and a tone stimulus to the other ear on each trial. His task was simply to write down which two stimuli he heard: (/ba/, /da/, or /ga/) and (high, medium, or low tone).

Results

The overall level of performance was excellent: the mean percent correct over all subjects and trials was 97 percent. The lowest identification score for any subject was 95 percent. This performance level is higher than those reported in the previous dichotic listening literature. Since there were virtually no errors, there was no opportunity for an ear advantage to occur: neither ear yielded superior performance, nor did it make any difference whether the stimuli were presented to the "proper" or "improper" ears.

In the present S/NS situation, then, we have two findings that conflict with the previous literature: essentially perfect identification accuracy and no ear effect.

Discussion

In light of these findings, perhaps we must reconsider what we mean by dichotic competition. Perceptual competition does not occur simply by presenting different messages to each ear at the same time, since S/NS pairs yielded virtually error-free performance. The S/NS condition does not appear to create a situation of information overload. The data fit the notion that there are two processors for auditory stimuli: a speech processor and a nonspeech processor. An ear advantage occurs only when a given processor is

Oscilloscope Photograph of a Sample Dichotic Trial Where Speech and Tone Began at the Same Time

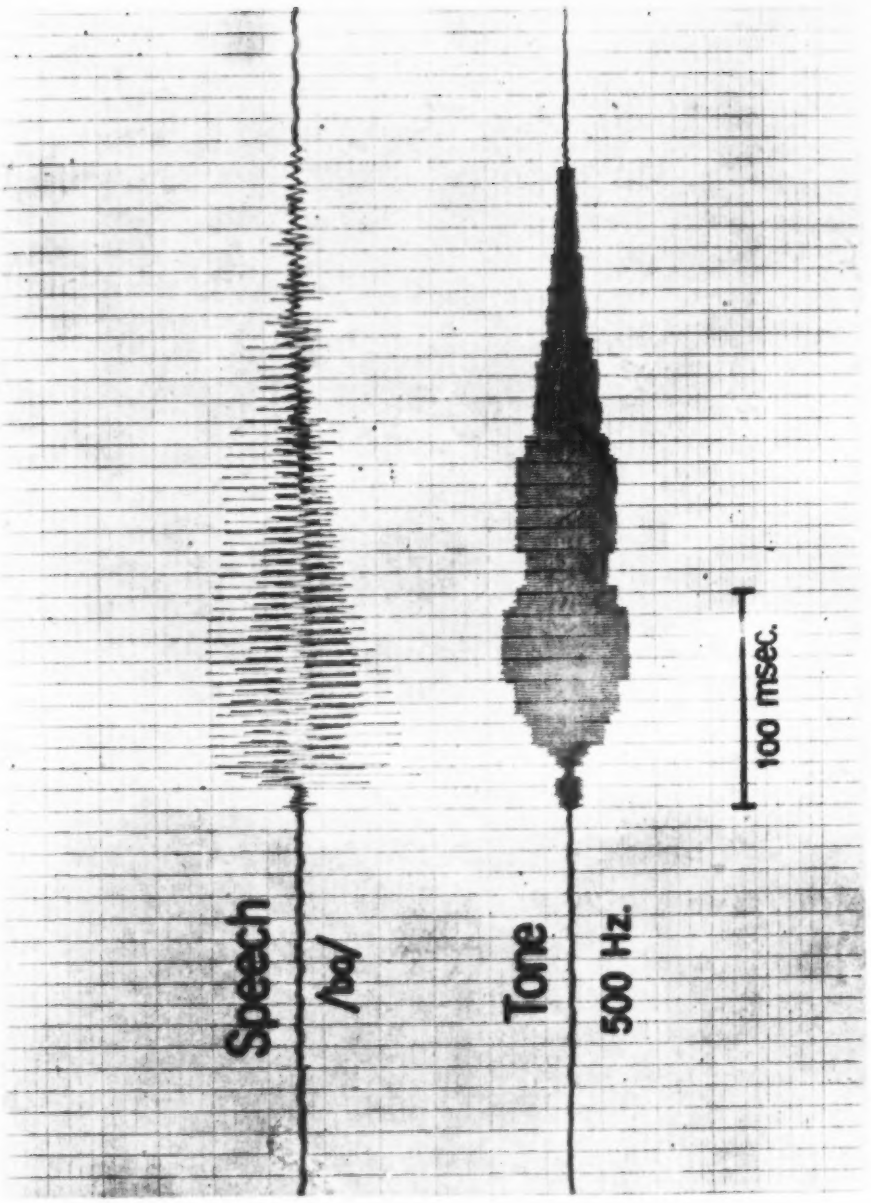


Fig. 2

overloaded, as when two speech stimuli are sent to the speech processor or when two nonspeech stimuli are sent to the nonspeech processor. In the present S/NS situation, the stimuli appear to be sent to different processors, each of which can perform its job without interference from the other.

But this is not the whole story. There was another task.

EXPERIMENT II - TEMPORAL ORDER JUDGMENT (TOJ) TASK

Method

Fifteen subjects (the ten from Exp. I plus five from the same pool) listened to the same stimulus tape. This time, each subject was asked to determine which stimulus began first on every trial, be it /ba/, /da/, /ga/, high, medium, or low tone. Therefore we are interested in percent correct temporal order judgment (TOJ) as a function of the lead-time conditions. Will performance again be equally accurate for both ears and will overall performance stay virtually error-free?

Results

First, let us examine the results for a single subject (MS) as shown in Figure 3. Here we have percent correct TOJ plotted as a function of ear and magnitude of lead. When the stimulus led in the left ear, performance was very good. This subject gave correct temporal order judgments over the entire range of left-ear leads. However, when the stimulus led in the right ear, performance was poor. The subject continued to report that the left-ear stimulus led. This effect was most striking at the short leads. Compare performance for the two ears at the 25-msec lead condition. MS was 94 percent correct for stimuli leading in the left ear and only 22 percent correct for stimuli leading in the right ear. Therefore, there was a net 72 percent advantage for the left ear over the right at 25 msec. This is a very large effect and contrasts with the typical ear advantage of approximately 10 percent reported in the previous literature (obtained for S/S identification trials). Similarly, at 50 msec, there was a net 55 percent left-ear advantage. Even at 75 msec, the left ear still had a 19 percent advantage. However, from 100 msec on outward, both ears performed equally well. Note that there is no "correct" response at the 0-lead point. Here we have simply entered the proportion of left- and right-ear responses, and attached the open circles to the appropriate sides of the continuum.

At the long leads MS performed very well; only at the short leads did he show an ear asymmetry. At what point did he judge the left-ear stimulus to lead 50 percent of the time and the right-ear stimulus 50 percent of the time--what we might call the "point of subjective equality"? The inputs for each dichotic trial have been aligned very precisely according to computer-controlled methods. So we know that the 0-lead items began at indeed the same point in time, down to 500 microsec accuracy. But it appears that MS does <u>not</u> know where the 0-lead cases are in the array of physical stimuli.

Sample Subject: Temporal Order Judgment (TOJ) Performance as a Function of Ear and Magnitude of Lead

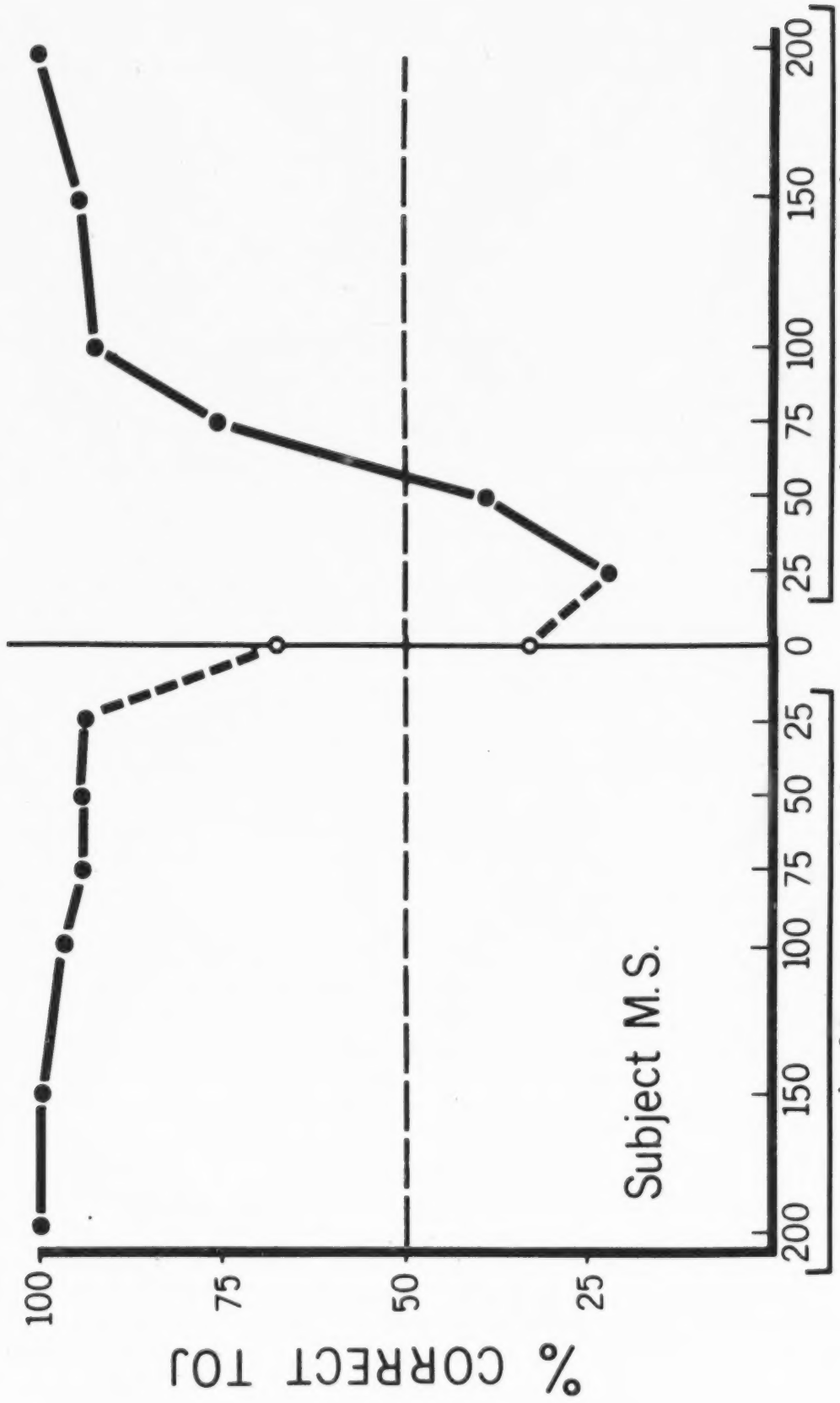


Fig. 3

In fact, the point of subjective equality appears to be out as far as about 60 msec for this subject. That is, the stimulus in the right ear must lead by at least 60 msec in order for him to respond at chance level.

We have seen this general pattern of results many times now, for most of the fifteen subjects in the present experiment and for about fifty others in variants of the task. Subjects differed in the magnitude of the left-ear advantage and in the place on the continuum where the right ear recovered. No subjects showed a significant right-ear advantage.

The results for all fifteen subjects are shown in Figure 4. The data are shown in histogram form to facilitate comparison of performance for the two ears. There is a left-ear advantage at each point on the continuum, with effects statistically significant out through the 100 msec lead condition. These pooled data represent 540 observations per ear for each lead case and 270 observations per ear for the 0-lead case. This display does no great disservice to any individual subject. In contrast with the identification task, then, the TOJ task yielded a large left-ear advantage and a decrease in overall level of performance, expecially at the shorter leads.

The left-ear advantage for the S/NS TOJ task is a very robust phenomenon. It appears to be immune to the effects of practice: performance remained stable over the course of the experiment, and subjects tested on several subsequent occasions gave comparable results in each session. None of the subjects was aware of the left-ear advantage when questioned in the postsession interview. When prompted to guess which ear yielded better performance, most guessed that they performed best on right-ear leads because they are right handed.

The effect carries with it some powerful phenomenological consequences. A brief anecdote concerning its discovery will illustrate this point. We were testing some new equipment designed to measure reaction times in another experiment. Briefly, the speech stimulus presented to the subject's earphones was also recorded on one channel of a response tape. The subject made one of two button-press responses on each trial. Each button activated an oscillator, and the resulting high or low tone was recorded on the second channel of the response tape. After a series of trials, the response tape was ready to be run through a data processor that would measure the time from onset of a given stimulus (speech) on one channel to the onset of the response (tone) on the other channel. One of the authors served as "subject" while the other monitored the response tape over dichotic earphones. On the first trial, the "monitor" heard a tone, then speech. The "subject" was asked to wait until he heard the onset of the speech stimulus before pressing the button. He replied that he had. Another trial was run. Again, the "monitor" heard tone-speech and warned the "subject" not to give false start. Again, the "subject" replied that he had heard the speech stimulus before he pressed the button. After a rather frustrating interchange, the "monitor" reversed the earphones and another trial was run. The effect went away. Under the original earphone configuration, the tone had gone to the left ear.

So far we have been discussing ear effects. What about stimulus effects? Subjects did show a bias or preference for reporting one class of stimuli over the other. For most, this bias was in favor of the speech stimulus. The stimulus effect may be related to labelling processes. The name for a speech

Group Data: Temporal Order Judgment (TOJ) Performance as a Function of Ear and Magnitude of Lead

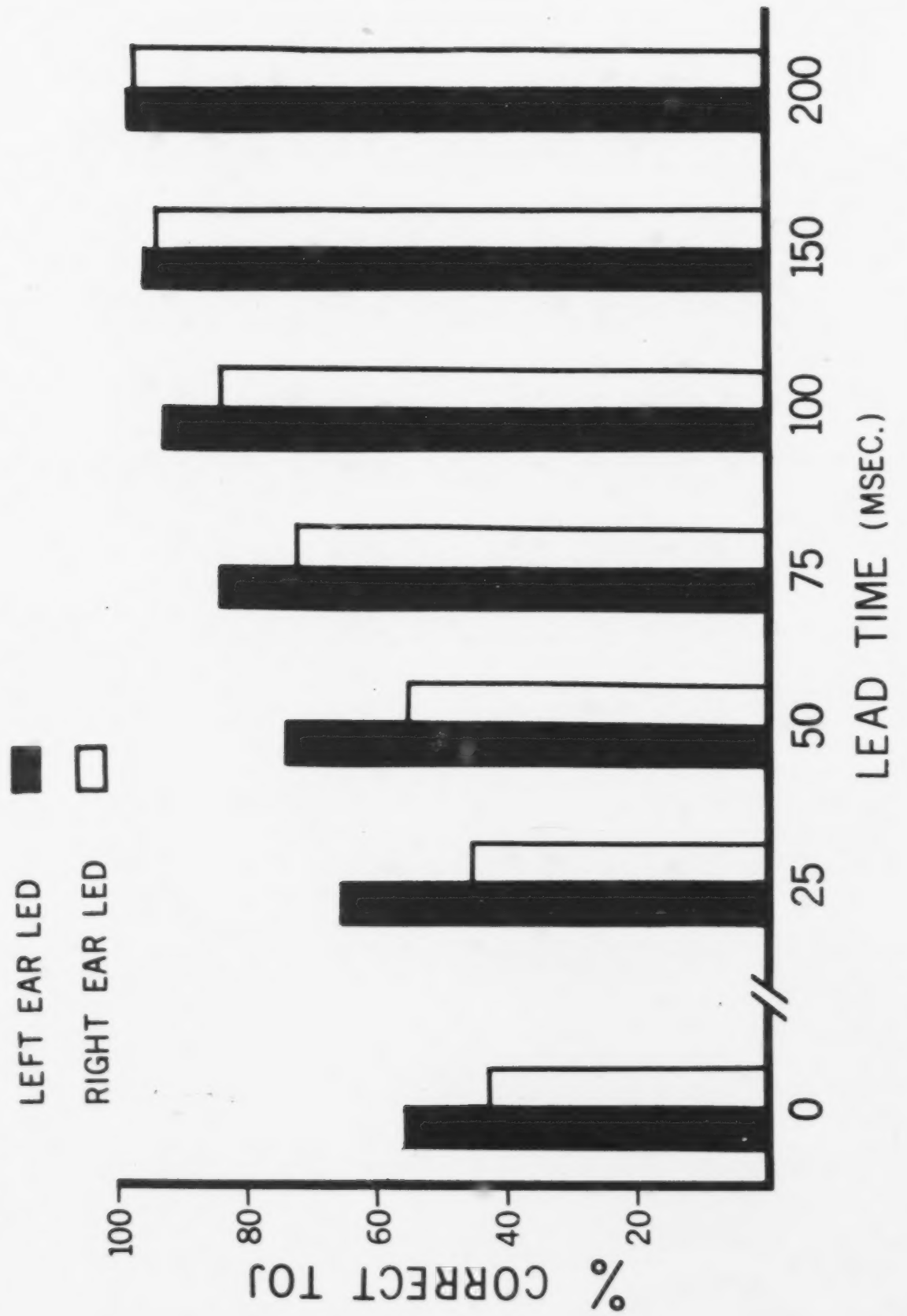


Fig. 4

stimulus is, in a sense, already in some sort of echoic memory before the subject makes his response, whereas he must search for the name to give to a tone stimulus, e.g., "medium" or "low." In any event, it is important to emphasize that subjects showed a wide range of stimulus preferences. On the other hand, the ear effect was unidirectional and was maintained no matter what type of preference an individual had for the stimuli.

Discussion

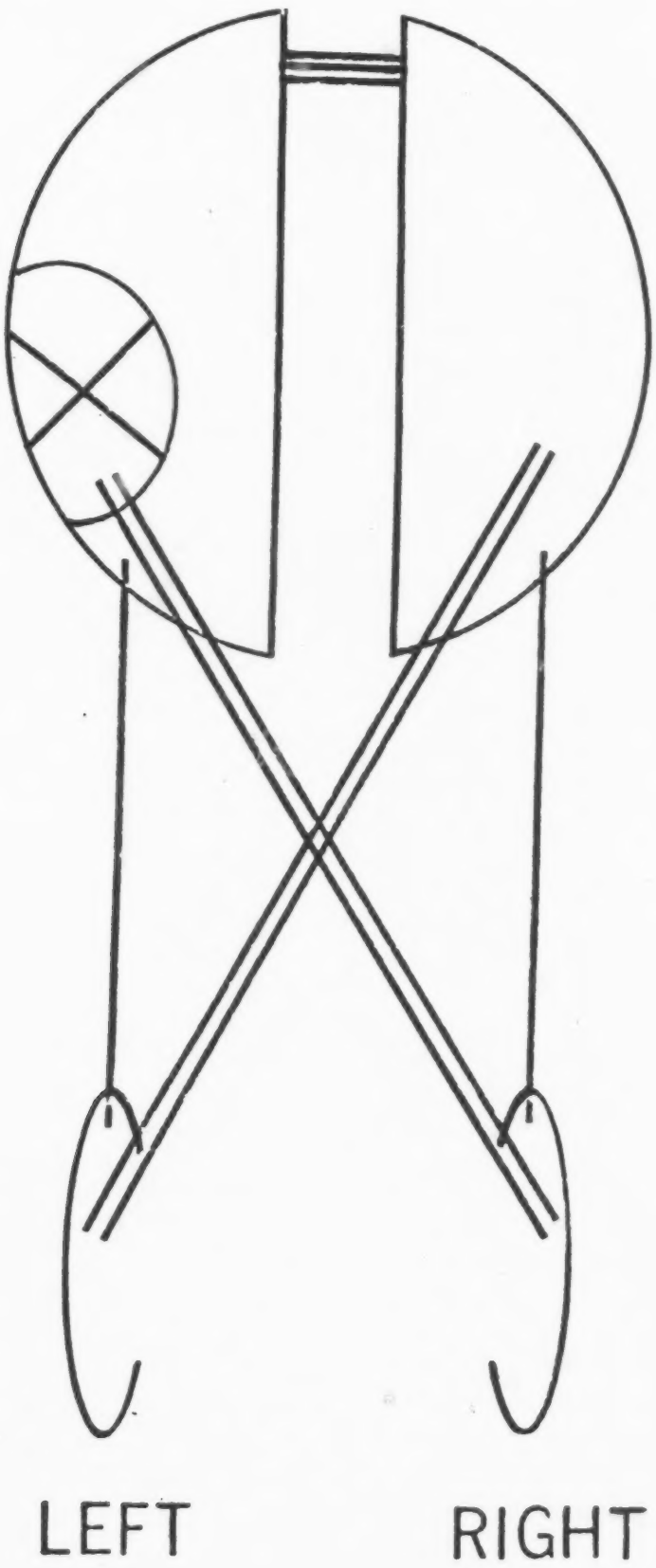
How are we to explain the left-ear advantage on the TOJ task? Unfortunately the present experiment does not provide enough data for a clear answer. Nevertheless, there are several approaches one might take to begin thinking about this robust and puzzling phenomenon.

Left-hemisphere function. Let us assume that the outlines of current dichotic listening models are reasonable (e.g., Kimura, 1967). As shown in Figure 5, we will assume 1) that the contralateral (crossed) connections from ears to hemispheres are prepotent in the dichotic listening situation, and 2) that the left hemisphere has a highly specialized processor, one that is designed to handle speech stimuli. If indeed the left hemisphere has some extra, very complicated machinery, then perhaps this machinery takes longer to "warm up" once stimuli are fed into it. Put crudely, the "turn around time" for the special processor might be on the order of 50-75 msec. However, this model needs amplification in order to deal with the entire dichotic listening literature. In the temporary absence of some appropriate follow-up data, discussion of the revised model would be merely speculative.

Right-hemisphere function. Many investigators have suggested that space is processed primarily in the right hemisphere (e.g., Carmon and Bechtoldt, 1969; Dorff et al., 1965; Kimura, 1963; Milner, 1958). The present S/NS data suggest that perhaps time is also processed more efficiently in the right hemisphere. Such a view would go counter to that of Efron (1965) who has suggested that time is handled back in the left hemisphere. But Efron was dealing with very simple situations such as light flashes and finger shock. Such situations may well involve different time perception mechanisms than those involved in speech. The view that time perception is handled primarily in a single hemisphere may well be an over-simplification. Day and Cutting (1971) recently found that TOJ accuracy yielded different ear advantages depending on the nature of the stimuli: S/S yielded a right-ear advantage while both NS/NS and S/NS yielded a left-ear advantage.

Modes of neural organization. So far we have been looking for specific functions that each hemisphere might perform: speech in the left hemisphere and time in the right hemisphere. There is an alternative approach with a different emphasis. Semmes (1968) has suggested that the two hemispheres represent contrasting modes of neural organization. The left hemisphere involves focal representation of functions, such as that for speech, whereas the right hemisphere involves diffuse representation of functions. To quote Semmes, "diffuse representation of elementary functions in the right hemisphere may lead to integration of dissimilar units." Therefore perhaps it is the comparison between dissimilar units in the S/NS TOJ task that produces the left-ear/right-hemisphere effect that we see so clearly in the data.

Schematic Diagram of the Connections from Ears to Hemispheres



44

Fig. 5

More work is needed in order to derive a satisfactory explanation for the S/NS data. No doubt various aspects of the above approaches will be helpful in solving this puzzle.

SUMMARY

On the identification task, there was no ear advantage, since performance was essentially perfect. On the TOJ task, there was a large left-ear advantage. Performance levels were near chance for 0-lead cases, and improved gradually to virtually perfect performance at the long leads.

There appear to be two necessary conditions for the large left-ear advantage to occur: 1) both speech and nonspeech must be put into the system, and 2) the subject must make a judgment that depends on relative time perception.

We have an interesting puzzle on our hands. We have a large and robust phenomenon that is not predicted by existing models of dichotic listening. Perhaps, then, the existing models only appear to explain the results now in the literature, and what is needed is a revised model that will handle both the old and the new phenomena.

REFERENCES

Borkowski, J. G., Spreen, O., and Stutz, J. Z. (1965) Ear preference and abstractness in dichotic listening. Psychonomic Science 3, 547-548.

Carmon, A. and Bechtoldt, H. P. (1969) Dominance of the right cerebral hemisphere for stereopsis. Neuropsychologia 7, 29-39.

Chaney, R. B., and Webster, J. C. (1966) Information in certain multidimensional sounds. J. Acoust. Soc. Amer. 40(2), 447-455.

Curry, F. K. W. (1967) A comparison of left-handed and right-handed subjects on verbal and non-verbal dichotic listening tasks. Cortex 3, 343-352.

Day, Ruth S. and Cutting, J. E. (1971) What constitutes perceptual competition in dichotic listening? Paper presented at the Eastern Psychological Association Meeting, New York, 17 April.

Dorff, J. E., Mirsky, A. F., and Mishkin, M. (1965) Effects of unilateral temporal lobe removals in man on tachistoscopic recognition in left and right visual fields. Neuropsychologia 3, 39-51.

Efron, R. (1965) The effect of handedness on the perception of simultaneity and temporal order. Brain 86, 261-284.

Kimura, D. (1961) Cerebral dominance and the perception of verbal stimuli. Canad. J. Psychol. 15(3), 166-171.

Kimura, D. (1963) Right temporal lobe damage. Arch. Neurol. 8, 264-271.

Kimura, D. (1964) Left-right differences in perception of melodies. Quart. J. Exper. Psychol. 16, 355-358.

Kimura, D. (1967) Functional asymmetry of brain in dichotic listening. Cortex 3, 163-178.

Mattingly, I., and Cooper, F. S. (1969) Computer-controlled PCM system. Paper presented at the Annual Meeting of the Acoustical Society of America, Philadelphia.

Milner, B. (1958) Psychological deficits produced by temporal-lobe excision. Proc. Assoc. Res. Nerv. Ment. Dis. 36, 244-257.

Semmes, J. (1968) Hemispheric specialization: A possible clue to mechanism. Neuropsychologia 5, 11-26.
Shankweiler, D., and Studdert-Kennedy, M. (1967) Identification of consonants and vowels presented to left and right ears. Quart. J. Exp. Psychol. 19(1), 59-63.

Temporal Order Perception of a Reversible Phoneme Cluster*

Ruth S. Day+
Haskins Laboratories, New Haven

Abstract. The syllable /t æ s/ was presented to one ear, while at the same time /t æ k/ was presented to the other ear. On some trials, both syllables began at the same time; on others, /t æ s/ led by 5, 10, 15...100 msec or /t æ k/ led by these same intervals. When asked to report "what they heard," subjects often reported /t æ sk/ or /t æ ks/. Later, when asked to report "the last sound they heard," subjects performed well on both /s/ and /k/. These results contrast with those of previous studies that used nonreversible clusters: when asked to report the first phoneme of the dichotic pair /bæŋkət/-/læŋkət/, subjects reported hearing /b/ first, independent of the lead conditions presented. A preliminary model that recognizes two levels of processing in temporal order perception is discussed.

At normal speaking rates, a listener can identify ten or more phonemes per second and perceive them in their correct temporal order. Yet the reversal of a single pair of phonemes can have some serious consequences. For example, compare the following two sentences: "I WANT YOU TO AXE HIM" and "I WANT YOU TO ASK HIM." How do we determine the sequential order of phonemes in the speech stream? Semantic and syntactic cues play important roles in this process in ordinary listening situations. However, when such higher-order variables are unavailable to the listener, how does he make temporal order judgments at the phoneme level?

Various methods for studying the perception of auditory sequences have been reported in the literature. However, it is difficult to apply the results of these experiments to the speech case since they used stimuli that were nonspeech (Hirsh, 1959) or at best "speech-like" (Broadbent and Ladefoged, 1959). Recently, Warren et al. (1969) have shown that perception of temporal order is very different for speech and nonspeech stimuli. Furthermore, the stimuli in most temporal order studies were presented in succession, either with or without a silent interval separating them. Complete successiveness rarely occurs in speech: each phoneme varies in its acoustic shape as a function of its neighbors such that at most points in time there is simultaneous information concerning two or more linguistic events.

---

*Paper presented at the seventy-ninth meeting of Acoustical Society of America, Atlantic City, 21-24 April 1970.

+Also, Yale University, New Haven.

Recent studies of phonemic fusion in dichotic listening (Day, 1970; Day, 1971; Day and Cutting, 1971) have approximated the speech perception situation more closely by presenting two speech stimuli at the same time with various relative onset times. The stimuli differed only in their initial phonemes, for example, /bæŋkət/ and /læŋkət/. On some trials, /bæŋkət/ led by 25, 50, 75, or 100 msec; on others, /læŋkət/ led by these same intervals, while on the remaining trials, both stimuli began at the same time. This technique enables information concerning two phonemic events (in this case /b/ and /l/) to be presented with varying degrees of simultaneity. The relative onset time parameter had surprisingly little effect on the perception of these items. When asked to report "what they heard" under dichotic stimulation, subjects were equally likely to report hearing the fused form /blæŋkət/ when /læŋkət/ led as when /bæŋkət/ led. On a different task, when specifically asked to judge the temporal order of the initial phonemes, most subjects were unable to do so. That is, they reported hearing /b/ first, no matter whether /b/ or /l/ actually led. Thus, instead of processing temporal order in an accurate fashion, subjects appeared to reflect the sequential dependencies of phonemes in the language. In English, (stop + liquid) clusters are permissible in initial position but (liquid + stop) clusters are not, i.e., items like */lbæŋkət/ cannot occur.

The results of these studies on phonemic fusion suggest that subjects are unable to perceive the correct order of phonemic events when there are temporal constraints imposed by the phonological rules of the language. What happens when these constraints are removed, namely when the to-be-fused cluster can occur in either order? Will processing of temporal order improve? The present study was designed to study perception of a reversible phoneme cluster.

EXPERIMENT I - IDENTIFICATION

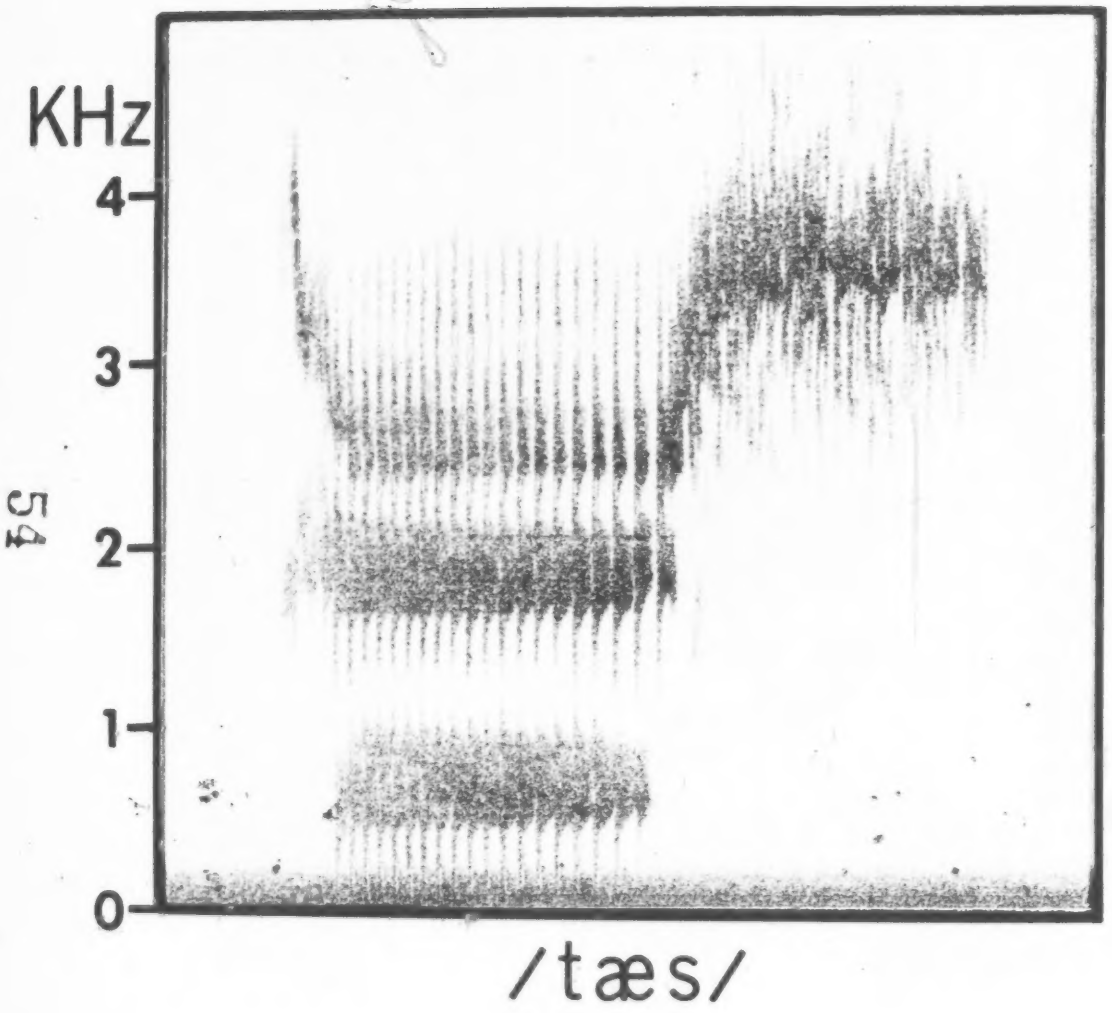
Method

Since there are no reversible clusters in initial position in English, a final cluster was selected. The stimuli were /tæs/ and /tæk/. Either order of the target phonemes is permissible in English: both /tæsk/ and /tæks/ are acceptable English words. The syllables were prepared on the computer-controlled parallel resonant synthesizer at the Haskins Laboratories and are displayed in Figure 1. (The fusion effect does not depend on the use of synthetic stimuli; several experiments have been performed using real speech, and substantial fusion levels have been obtained in all cases.) Formant frequencies, amplitudes, and other parameters were identical for the /tæ-/ portions of both stimuli. Thus they differed only in their final phonemes, /s/ and /k/.

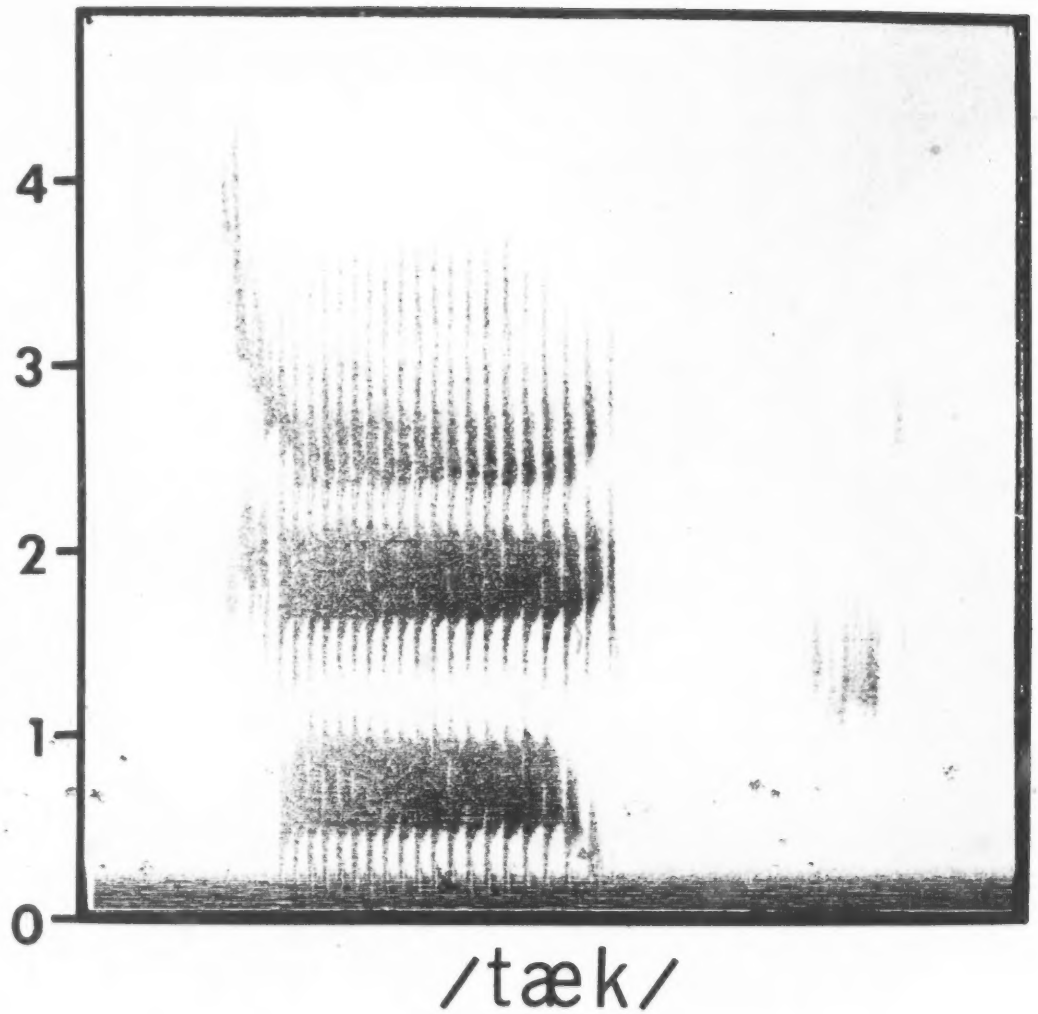
All trials were dichotic pairs composed of /tæs/ to one ear and /tæk/ to the other ear. The onsets of the syllables were aligned over a wide range of values: on some trials, /tæs/ led in steps of 5 msec out to a 100 msec lead; on other trials, /tæk/ led by these same values; and on the remaining trials, both inputs began at the same time.

Each of twenty right-handed subjects was asked to report "whatever he heard" on every trial and was informed that the trials might consist of one

Spectrograms of the Stimuli Used in Dichotic Presentation



/tæs/



/tæk/

Fig. 1

word or two, real words or nonsense words. No mention was made of the dichotic technique. All of the appropriate counterbalancing procedures were observed.

Results

Fusion level. The stimuli fused readily. About three-fourths of all responses were fusions.

Effect of lead time on fusion level. Figure 2 plots the probability of a fusion response as a function of the various lead time conditions. Lead time had relatively little effect on fusion level. Fusions occurred readily, both when /tæs/ led and when /tæk/ led. Furthermore, the magnitude of a given item's lead had little effect on fusion level.

"Correct" phonemic order. In computing the fusion probabilities shown in Figure 2, either /tæsk/ or /tæks/ was scored as a fusion, regardless of the lead arrangements. No attempt was made to determine whether the fusions reflected the order in which /s/ and /k/ were presented. Figure 3 takes this consideration into account. When /tæs/ led the "correct" phonemic order was /tæsk/; similarly, when /tæk/ led, the "correct" phonemic order was /tæks/. Scores were unequal on the two sides of the display, with /tæk/-lead items yielding a higher level of "correct" phonemic fusions. Since there can be no "correct" phonemic order at the 0-lead case, responses of both types have been plotted separately and attached to the appropriate side of the continuum: 80 percent of all responses here were /tæks/ while 20 percent were /tæsk/.

Discussion

How can we account for these findings? The answer appears to lie below the phonemic level. A variety of evidence suggests that the acoustic shapes of stop consonants undergo greater changes as a function of context than do other classes of phonemes, such as fricatives. If so, then the particular acoustic (or phonetic) shape of the /k/ in /tæk/ is more important than that of the /s/ in /tæs/ in that it may bias the perceived order of the two phonemes. This situation is illustrated in Figure 4. Consider the phonetic form of the /k/ in /tæk/. The argument is that this /k/ is more like a /k/ that could be followed by an /s/ than one that could be preceded by an /s/. Therefore, given that a /k/ is to be fused, it is more likely to be heard as /tæks/ than as /tæsk/. On the other hand, the /s/ of /tæs/ is similar both to an /s/ that could be followed by a /k/ and to one that could be preceded by a /k/. Therefore, given that an /s/ is to be fused, it is equally likely to be heard as /tæsk/ or /tæks/.

The appeal here is being made to the degree of acoustic variation for different phoneme classes. It is argued that the acoustic parameters important for the perception of stops undergo more change as a function of adjacent phonemes than do the acoustic parameters important for the perception of fricatives. These notions can be tested by synthesizing alternative forms of the target phonemes and observing the resulting fusion levels. For example, when the acoustic shape of /k/ is more like that of a /k/ that could be preceded by an /s/, will fusions then shift toward the /tæsk/ response? This type of experiment may prove fruitful for determining what the acoustic cues for cluster perception might be.

Fusion Response Probability as a Function of Stimulus and Magnitude of Lead

Fig. 2

Conditional Probability of "Correct" Phonemic Order on Fusion Trials

Fig. 3

Phonetic Form of Target Phonemes Relative to Their Forms in a Reversible Cluster

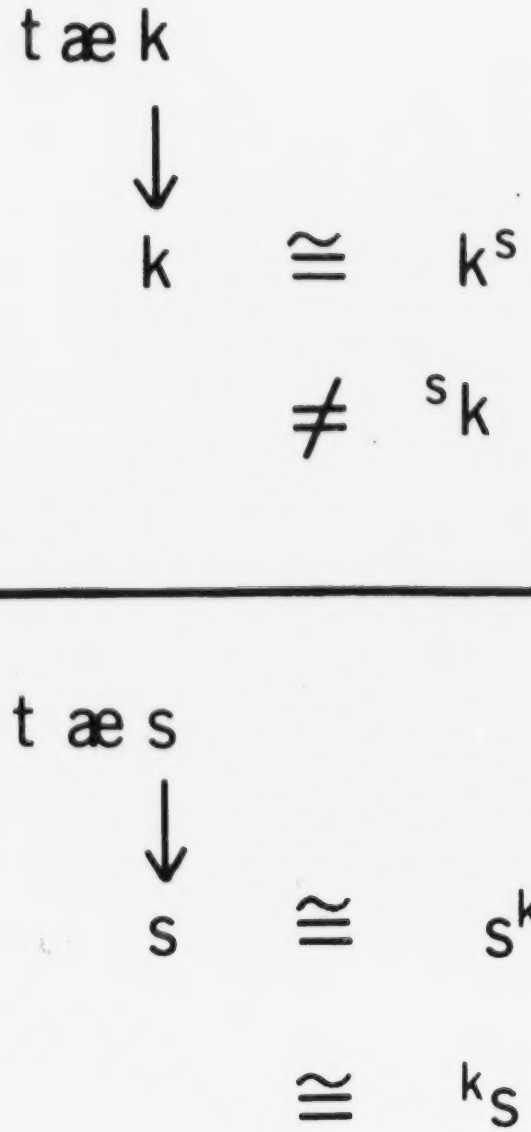


Fig. 4

EXPERIMENT II - TEMPORAL ORDER JUDGMENT (TOJ)

Method

A second task was given to see whether the same subjects could determine the temporal order of the target phonemes when specifically asked to do so. On each trial, the subject simply reported "the last sound he heard."

Results

TOJ accuracy. Performance was very good for both /s/ and /k/. These results contrast with those obtained in the nonreversible cluster experiments, as shown in Figure 5. The left side of the display shows that overall performance on nonreversible clusters was poor: when the stop led, subjects reported hearing the stop first, but when the liquid led, they still reported hearing the stop first. Thus there was a failure to discriminate between the two stimulus situations. Subjects reported hearing the stop first, independent of the stimulus events. When the cluster was reversible, as in the present experiment, performance improved substantially. Subjects were able to report correct temporal arrangements for either presentation order. It should be pointed out, however, that performance was somewhat better for one of the stimulus orders, namely, for items where /k/ led and /s/ was last. These are the same items that yielded more "correct" phonemic fusions in the identification task. Hence these TOJ findings are compatible with the fusion data.

In contrast with the nonreversible case, temporal order judgment was very good when the cluster could occur in either order in the language. Nevertheless, one of the temporal orders was somewhat more preferred.

Discussion

How, then, do we determine the temporal order of phonemic events in the speech stream? A simple time-tagging model cannot handle the present results. Such a view suggests that, as each phoneme arrives, it gets "tagged" with a label that records its time of entry. Thus, the first phoneme arrived at time t_1, the second at t_2, and the nth at t_n. When asked to report the ith item, the subject simply searches through the time tags and selects the phoneme tagged t_i. This model would predict comparable temporal order performance over all types of phoneme pairs.

A model that recognizes two levels of processing is better able to handle the data. When the target phonemes can occur in only one order, as in the /bæŋkət/-/læŋkət/ case, linguistic rules gain the upper hand: most subjects can only hear the phonemes in the order permitted by the phonological rules. Nonlinguistic considerations such as actual temporal order or acoustic shape of the target phonemes do not influence perception. However, when the phonological constraints are removed, as in the /tæs/-/tæk/ case, subjects can perform successful temporal analyses. The fact that one of the temporal orders is somewhat more preferred suggests that acoustic shape does influence perception even in this situation.

There appear to be two general levels of processing: a linguistic level and a nonlinguistic level. Evidence that these two levels can be separated comes from some recent studies (Day, 1971; Day and Cutting, 1971) of nonreversible clusters. As expected, subjects reported hearing the stops first,

Comparison of Temporal Order Judgment (TOJ) Performance on Reversible and Nonreversible Clusters

NON-REVERSIBLE CLUSTERS

(pooled over 5 experiments)

STIMULI \ RESPONSES	stop led	liquid led
stop led	.75	.25
liquid led	.70	.30

REVERSIBLE CLUSTERS

(present experiment)

STIMULI \ RESPONSES	/s/ last	/k/ last
/s/ last	.70	.30
/k/ last	.39	.61

Fig. 5

independent of the lead conditions. However, when asked to judge temporal order at a nonlinguistic level, namely by reporting which ear led, performance was equally good for both liquid-leading and stop-leading items.

Both linguistic and nonlinguistic processing levels operate in normal listening situations. But the linguistic level appears to be prepotent: it can effect selective loss of information obtained from the nonlinguistic level. Correct temporal order may be represented in the system at some point in time, but later stages of processing mold this information to conform to the linguistic matrix of the language. Hence nonlinguistic information, concerning acoustic shape and temporal order information, may be lost or ignored.

It may be that even higher-order variables concerned with semantic and syntactic context can override temporal order information obtained at the phonemic level. Given a dichotic pair that is usually fused in a particular order, e.g., /tæks/, will perception change when the item is embedded in a context that demands the reverse phonemic order, e.g., "CHOPPING WOOD IS A VERY HARD _____"?

In conclusion, the value of proposing two levels of processing is that it allows one to investigate at what point information is lost, or disregarded. The view here is that temporal order information is lost only after it enters higher stages of linguistic processing.

REFERENCES

Broadbent, D.E. and Ladefoged, P. (1959) Auditory perception of temporal order. J. Acoust. Soc. Amer. 31, 1539.

Day, Ruth S. (1970) Temporal order judgments in speech: Are individuals language-bound or stimulus-bound? Haskins Laboratories Status Report, SR-21/22, 71-87.

Day, Ruth S. (in press) Release from language-bound perception. Haskins Laboratories Status Report, SR-25/26.

Day, Ruth S. and Cutting, James E. (1970) Levels of processing in speech perception. Paper presented at the Tenth Annual Meeting of the Psychonomic Society, San Antonio, November.

Hirsh, I.J. (1959) Auditory perception of temporal order. J. Acoust. Soc. Amer. 31, 759-767.

Warren, R.M., Obusek, C.J., Farmer, R.M., and Warren, R.P. (1969) Auditory sequence: Confusion of patterns other than speech or music. Science 164, 586-587.

Phonetic Ability and Related Anatomy of the Newborn and Adult Human, Neanderthal Man, and the Chimpanzee

Philip Lieberman,[+] Edmund S. Crelin,[++] and Dennis H. Klatt[+++]

Human language is one of the defining characteristics of modern man. Although the evolution of human language has been the subject of hundreds of books and essays,[1] not much is presently known. In recent years the primary focus has been directed towards the nature of the mental ability that may underlie the syntactic and semantic aspects of human language. The phonetic aspect of human language has been neglected. This follows from a rather common opinion concerning language, i.e., that its phonetic aspect is trivial and indeed finally irrelevant to the serious study of human language and its evolution. Simpson (1966), for example, reviewing attempts to trace the evolution of language, notes that

> audible signals capable of expressing language do not require any particular phonetic apparatus, but only the ability to produce sound, any sound at all. Almost all mammals and a great number of other animals can do that. Moreover, a number of animals, not only birds but also some mammals, can produce sounds recognizably similar to those of human language, and yet their jaws and palates are radically nonhuman.

Simpson essentially sets forth two premises. First, that any arbitrary set of sounds would serve as a phonetic base for human language. Second, that many animals also can produce the sounds that, in fact, occur in human

---

[+]Haskins Laboratories, New Haven, and the University of Connecticut, Storrs.

[++]Department of Anatomy, Yale University School of Medicine, New Haven.

[+++]Massachusetts Institute of Technology and Research Laboratory of Electronics, Cambridge.

Acknowledgment: We would like to thank Dr. E.L. Simons of Yale University for the chimpanzee specimen, Dr. P.F. Marler of Rockefeller University for making tape recordings and spectrograms of chimpanzee utterances available, and Drs. A.M. Liberman and C. Darwin of Haskins Laboratories and Dr. W.S. Laughlin of the University of Connecticut for their many useful comments. This work was supported, in part, by U.S. Public Health Service Grants AM-09499, HD-01994, DE-01774, and NB-04332-8.

[1]Hewes (1971) has compiled a comprehensive annotated bibliography on the evolution of language.

language. If Simpson's premises were true there would be little point in attempting to trace the evolution of human linguistic ability by studying either the comparative phonetic abilities of modern man and other living animals, or in attempting to reconstruct the phonetic abilities of extinct fossil hominids from their skeletal remains. Neither premise, however, is true. The results of research on the perception of human speech have shown that human language depends on the existence of the particular sounds of human speech. No other sounds will do. The results of recent research on the anatomic basis of human speech have likewise demonstrated that no living animal, other than modern man, has the vocal mechanism that is necessary to produce the sounds of human speech.

We have discussed some of the anatomical factors that prevent living nonhuman primates and newborn humans from producing the range of sounds that characterize human speech (Lieberman, 1968, 1969; Lieberman et al., 1968, 1969).[2] We have also been able to reconstruct the vocal apparatus of "classic" Neanderthal man (Lieberman and Crelin, 1971). Our present paper has two objectives. We shall compare the anatomy and speech-producing ability of the vocal mechanism of adult modern man with adult chimpanzee, newborn modern man, and the reconstructed vocal mechanism of adult "classic" Neanderthal man. We will then discuss the speech-perceiving and general linguistic abilities of chimpanzee and Neanderthal man in the light of their sound-making abilities. We shall, in this regard, consider some recent theoretical and experimental studies that relate the production and the perception of speech.

ACOUSTIC THEORY OF SPEECH PRODUCTION

The acoustic theory of speech production (Chiba and Kajiyama, 1958; Fant, 1960) relates the vocal mechanism to the acoustic signal. Human speech essentially involves the generation of sound by the mechanism of vocal-cord vibration and/or air turbulence and the acoustic shaping of these sound sources by the resonances of the supralaryngeal vocal tract. The shape of the human supralaryngeal vocal tract continually changes during the production of speech. These changes in the supralaryngeal vocal tract change its resonant properties. A useful mechanical analog to the aspect of speech production that is of concern to this discussion is a pipe organ. The musical function of each pipe is determined by its length and shape. (The pipes have different lengths and may be open at one end or closed at both ends.) The pipes are all excited by the same source. The resonant modes of each pipe determine the note's acoustic character. In human speech the phonetic properties that differentiate vowels like /i/ and /a/ from each other

---

[2]These results are consistent with the fact that it has never been possible to train a nonhuman primate to talk. Kellogg (1968) reviews a number of recent attempts at raising chimpanzees as though they were children. It is interesting to note that similar attempts date back to at least the eighteenth century (La Mettrie, 1747). The "speech" of "talking birds" is not similar to human speech at the acoustic or anatomic levels (Greenewalt, 1967). A parrot's imitation of human speech is similar to a human's imitation of a siren. The signal is accepted as a mimicry. It has different acoustic properties than the siren's signal, and it is produced by a different apparatus.

are determined by the resonant modes of the supralaryngeal vocal tract. The frequencies at which resonances occur are called "formant" frequencies.

The acoustic theory of speech production which we have briefly outlined thus relates an acoustic signal to a supralaryngeal vocal-tract configuration and a source. It is therefore possible to determine some of the constraints of an animal's phonetic range if the range of supralaryngeal vocal-tract variation is known. The phonetic repertoire of an animal can obviously be expanded if different sources are used with similar supralaryngeal vocal-tract configurations. We can, however, isolate the constraints that the range of supralaryngeal vocal-tract variation will impose on the phonetic repertoire.

VOCAL-TRACT ANATOMY

The anatomic specializations of modern man that are necessary for human speech are evident when we compare the supralaryngeal vocal tract of adult man with creatures who lack human speech. We will start with a brief account of the skeletal similarities among Neanderthal man and newborn modern man[3] and adult chimpanzee that make it possible to reconstruct the supralaryngeal vocal tract of Neanderthal man.

---

[3]The similarity between human newborn and the adult Neanderthal fossil conforms to the view that modern man and Neanderthal man had a common ancestor. Darwin in On the Origin of Species (1859, p. 449) clearly states the premise that we are following in making this inference. He states that "In two groups of animals, however much they may at present differ from each other in structure and in habits, if they pass through the same or similar embryonic stages, we may feel assured that they have both descended from the same or nearly similar parents, and are therefore in that degree closely related." The adult Neanderthal skull has certain specialized features, like a supraorbital torus, that are not present in newborn modern man nor in adult modern man. This indicates that Neanderthal man is probably not directly related to modern man. He is, as Boule (1911-1913) recognized, probably an early offshoot from the mainstream of hominids that evolved into modern man. The skulls of present-day newborn apes are quite similar to the human newborn (Schultz, 1968). This would indicate an early common ancestral form for both present-day apes and man. It does not show that modern man has evolved by retaining infantile characteristics. Adult modern man, in his own way, deviates as much from his newborn state (Crelin, 1969; Lieberman and Crelin, 1971) as adult living apes do from their newborn form.

Physical anthropologists and anatomists have noted, over the years, that measurements of particular aspects of Neanderthal skulls fall within the range of variation that may be found in modern man (Patte, 1955). This finding is not surprising since all adult modern men develop from the newborn morphology which has many similarities to that of adult "classic" Neanderthal man. The course of human maturation is not even and some individuals fail to develop "normally." In extreme pathologic conditions like Down's Syndrome, the individual may, in fact, retain many aspects of the newborn morphology, especially those of the skull. Benda (1969) notes that Down's Syndrome may be characterized, in part, as a developmental problem. We have examined a number of subjects afflicted with Down's Syndrome who cannot produce "articulate"

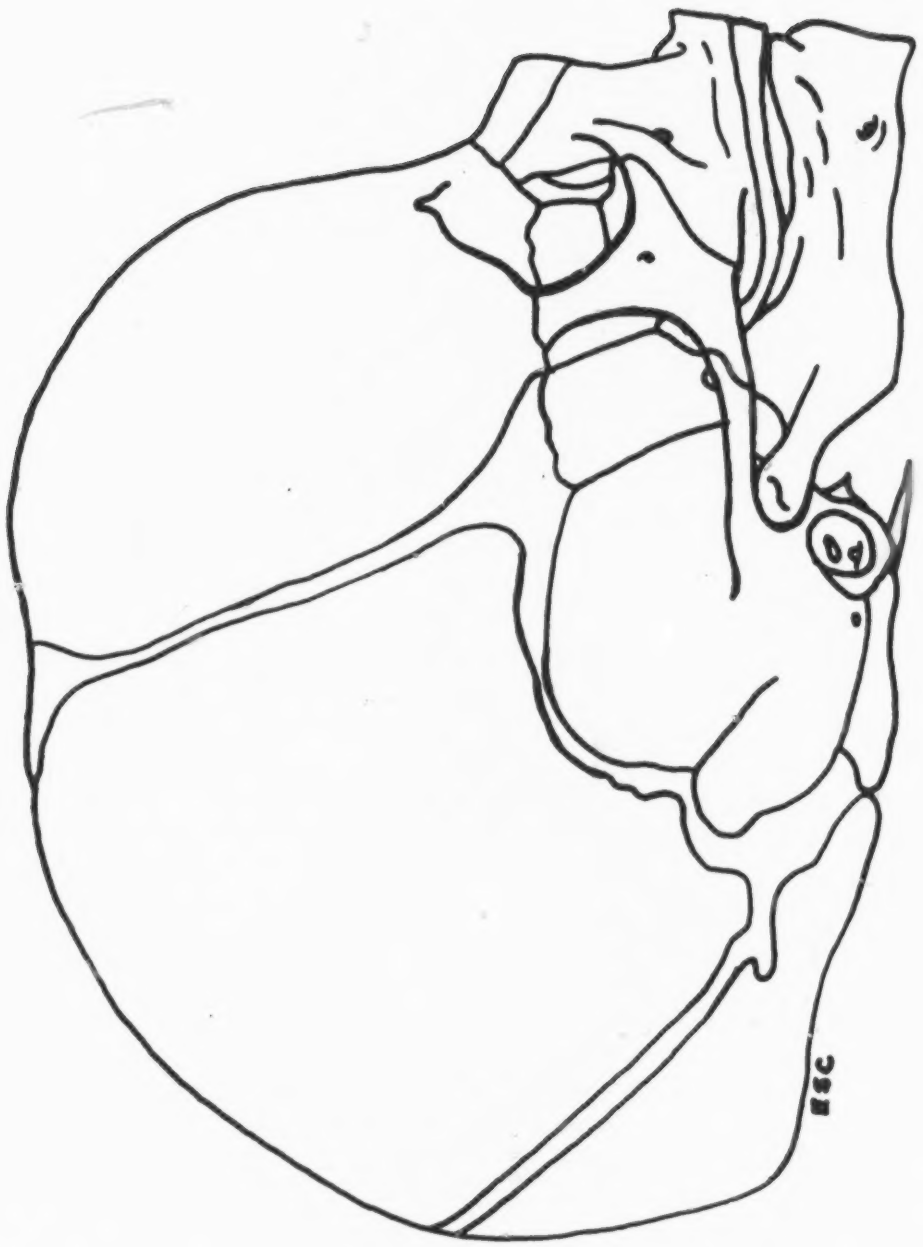
In Figures 1 through 4 lateral views of the skulls of newborn man, adult chimpanzee, the La Chapelle-aux-Saints Neanderthal man, and adult modern man are presented. The skulls have all been drawn to appear nearly equal in size. Skull features of newborn, chimpanzee, and Neanderthal man that are similar to each other, but different from that of adult modern man, are as follows: (1) they have a generally flattened out skull base; (2) they lack mastoid processes (very small in Neanderthal); (3) they lack a chin (occasionally present in newborn); (4) the body of the mandible is much longer than the ramus (about 60 to 100 percent longer); (5) the posterior border of the mandibular ramus is markedly slanted away from the vertical plane; (6) the mandibular foramen leading to the mandibular canal has a more horizontal inclination; (7) the pterygoid process of the spheroid bone is relatively short and its lateral lamina is more inclined away from the vertical plane; (8) the styloid process is more inclined away from the vertical plane; (9) the dental arch of the maxilla is U-shaped instead of V-shaped; (10) the basilar part of the occipital bone between the foramen magnum and the sphenoid bone is only slightly inclined away from the horizontal toward the vertical plane; (11) the roof of the nasopharynx is a relatively shallow elongated arch; (12) the vomer bone is relatively short in its vertical height and its posterior border is inclined away from the vertical plane; (13) the vomer bone is relatively far removed from the junction of the sphenoid bone and the basilar part of the occipital bone; (14) the occipital condyles are relatively small and elongated.

The chimpanzee differs from newborn and adult modern man and Neanderthal man insofar as its mandible has a "simian shelf," i.e., internal buttressing of the anterior portion of mandible. The simian shelf inhibits the formation of a large air cavity behind the teeth. In adult man a large cavity behind the teeth can be formed by pulling the tongue back in the mouth.

---

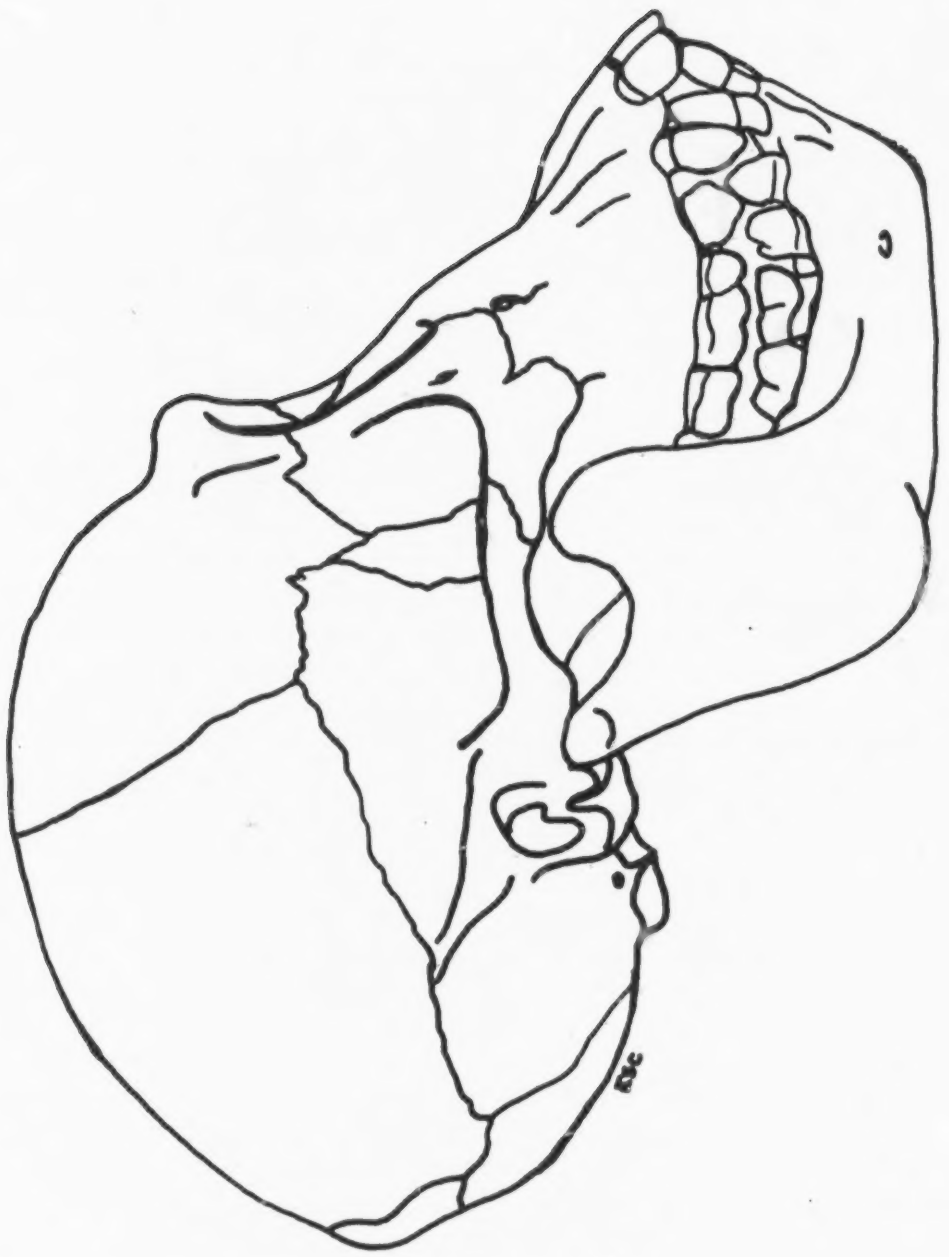
speech (Lieberman and Crelin, unpublished data). Some of these subjects may lack the mental ability that is necessary to control their vocal apparatus, but some of them appear to have vocal tracts that resemble the normal newborn vocal tract. They, in effect, have Neanderthaloid vocal tracts, and they cannot produce human speech. The base of their skulls and their mandibles generally resemble those of a Neanderthal. It is therefore not surprising that Virchow (1872) believed that the original Neanderthal skull, which was found in 1856, was either a pathologic specimen or the skull of an imbecile.

It is also evident that different population groups of modern man have somewhat different skeletal features. In some population groups a particular skeletal feature will fall within the range characteristic of classic Neanderthal man. Laughlin (1963), for example, notes that the breadth of the ramus of the mandible in Eskimos and Aleuts can exceed the breadth of this feature in Neanderthal man. The length of the body of the mandible is also somewhat longer for Aleuts and Eskimos than is the case for other modern human skulls. The length of the body of the mandible can be about 20 percent greater than the ramus in an adult male Aleut skull. This value is, however, much smaller than is the case for either Neanderthal man or newborn human where the length of the body of the mandible is 60 to 100 percent greater than the ramus (as measured on a lateral projection to the midline of the mandible). The total ensemble of skeletal features of the base of the skull for Aleuts and Eskimos is, moreover, consistent with the "angulation" of the vocal tract of adult modern man.



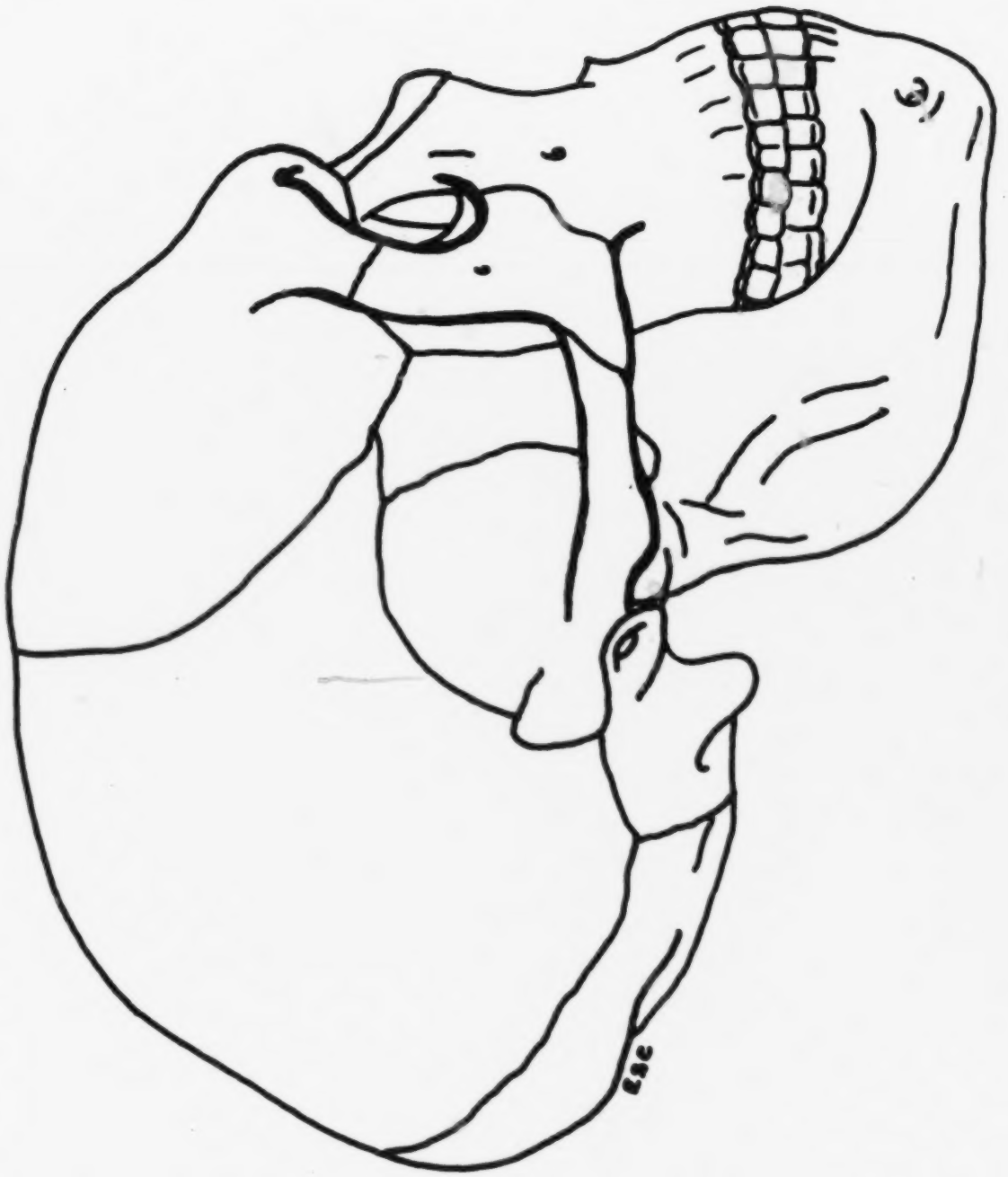
SKULL OF A HUMAN NEWBORN

Fig. I



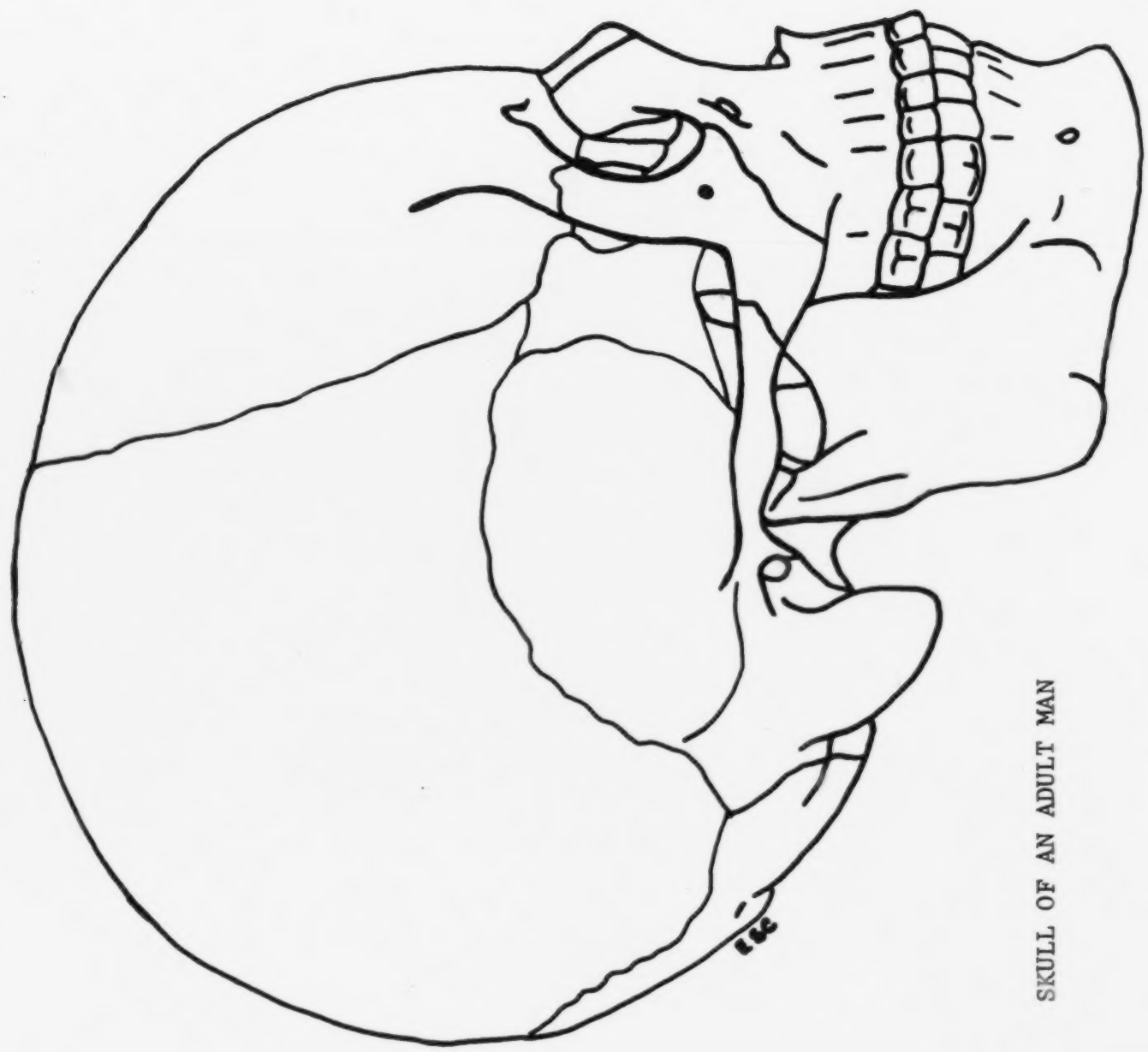
SKULL OF AN ADULT CHIMPANZEE

Fig. 2



SKULL OF THE LA CHAPELLE-AUX-SAINTS FOSSIL NEANDERTHAL MAN

Fig. 3



SKULL OF AN ADULT MAN

Fig. 4

The significance of these skeletal features can be seen when the supralaryngeal vocal tracts that correspond to these skulls are examined. The chimpanzee specimen used in this study was the head and neck of a young adult male sectioned in the midsagittal plane (Figure 5). The human newborn and adult specimens were those described by Lieberman and Crelin (1971) which included a number of heads divided in the midsagittal plane. Silicone-rubber casts were made of the air passages, including the nasal cavity, of chimpanzee, newborn, and adult man. This was done by filling each side of the split air passages separately in the sectioned heads and necks to insure perfect filling of the cavities. The casts from each side of a head and neck were then fused together to make a complete cast of the air passages. The cast of the Neanderthal air passages was made from the reconstructed nasal, oral, pharyngeal, and laryngeal cavities of the La Chapelle-aux-Saints fossil (Lieberman and Crelin, 1971). All four casts are shown in the photograph in Figure 6.

Even though the cast of the newborn air passages is much smaller than those of chimpanzee and adult modern man and Neanderthal man, it is apparent that the casts of newborn and chimpanzee are quite similar. When outlines of the air passages from all four are made nearly equal in size in Figure 7, one can more readily recognize what the basic differences and similarities are:

(1) Newborn human, chimpanzee, and Neanderthal man all have their tongue at rest completely within the oral cavity, whereas in adult man the posterior third of the tongue is in a vertical position forming the anterior wall of the supralaryngeal pharyngeal cavity. The foramen cecum of the tongue is thus located far more anteriorly in the oral cavity in chimpanzee and newborn than it is in adult man.

(2) In newborn, chimpanzee, and Neanderthal the soft palate and epiglottis can be approximated, whereas they are widely separated in adult man and cannot approximate.

(3) There is practically no supralaryngeal portion of the pharynx present in the direct airway out from the larynx when the soft palate shuts off the nasal cavity in chimpanzee, Neanderthal, and newborn man. In adult man half of the supralaryngeal vocal tract is formed by the pharyngeal cavity. This difference between chimpanzee, Neanderthal, and newborn and adult man is a consequence of the opening of the larynx into the pharynx, which is immediately behind the oral cavity in chimpanzee, Neanderthal, and newborn. In adult man this opening occurs farther down in the pharynx. Note that the supralaryngeal pharynx in adult man serves both as a pathway for the ingestion of food and liquids and as an airway to the larynx. In chimpanzee, Neanderthal, and newborn man the section of the pharynx that is behind the oral cavity is reserved for deglutition. The high epiglottis can, moreover, close the oral cavity to retain solids and liquids and allow unhampered respiration through the nose.

(4) The level of the vocal folds (cords) at rest in chimpanzee is at the upper border of the fourth cervical vertebra, whereas in adult man it is between the fifth and sixth in a relatively longer neck. The position of the hyoid bone is high in chimpanzee, Neanderthal, and newborn. This is concomitant with the high position of the larynx.

LEFT HALF OF THE HEAD AND NECK OF A YOUNG ADULT MALE CHIMPANZEE
SECTIONED IN THE MIDSAGGITAL PLANE

Fig. 5

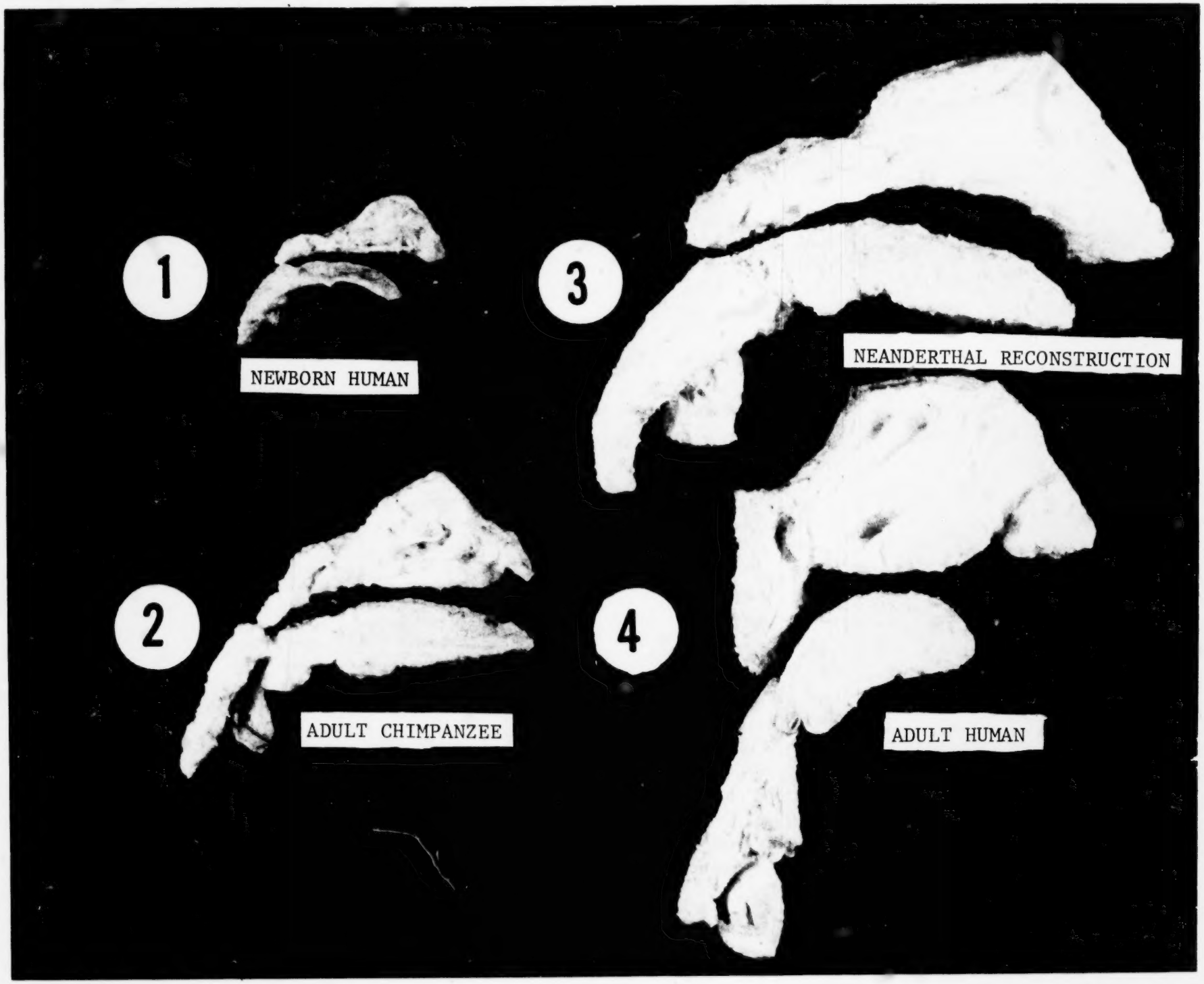


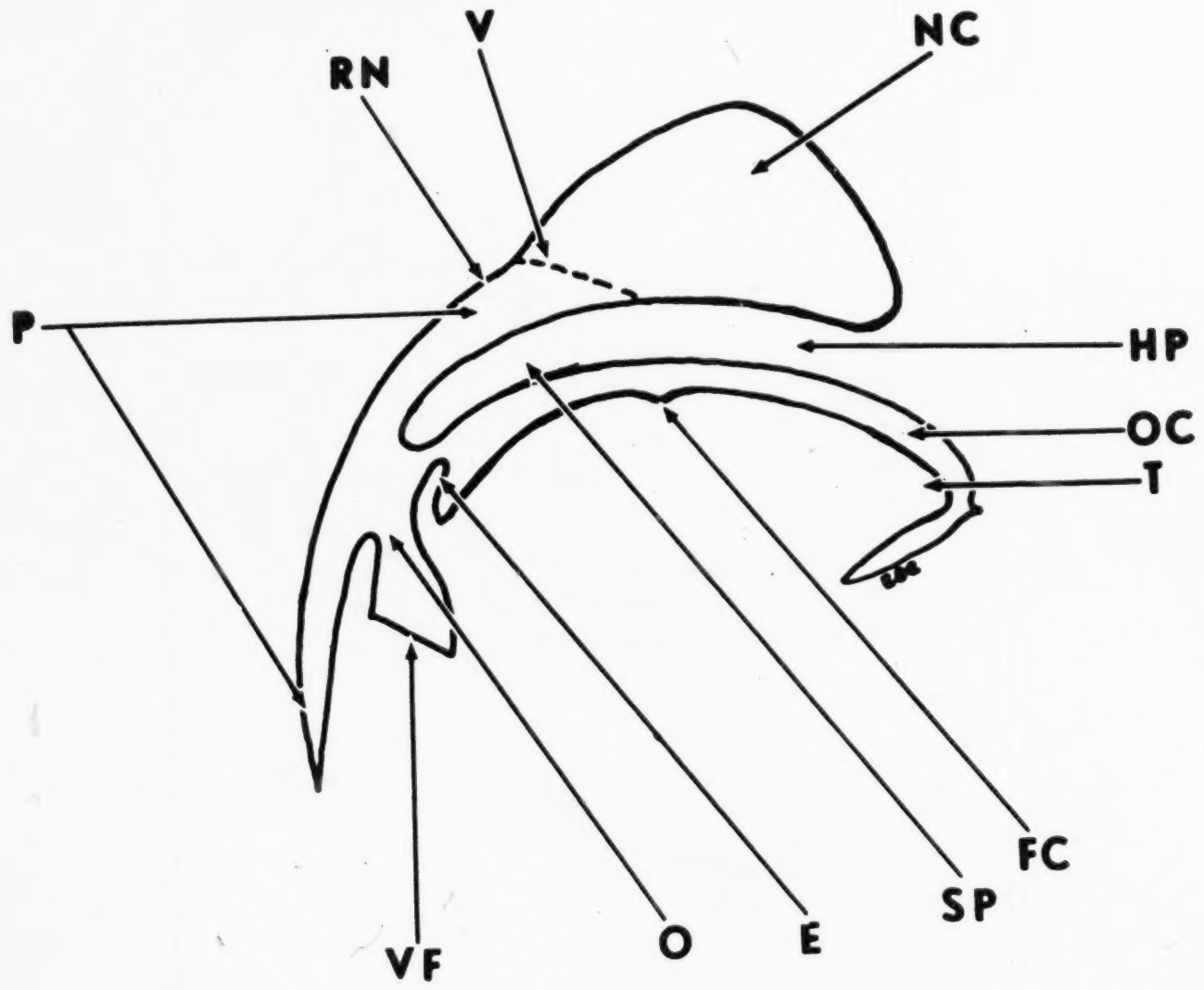
Fig. 6

CASTS OF THE NASAL, ORAL, PHARYNGEAL, AND LARYNGEAL CAVITIES

DIAGRAM OF THE AIR PASSAGES OF NEWBORN HUMAN

Fig. 7a

DIAGRAM OF THE AIR PASSAGES OF ADULT CHIMPANZEE



P - Pharynx
RN - Roof of Nasopharynx
V - Vomer Bone
NC - Nasal Cavity
HP - Hard Palate
OC - Oral Cavity
T - Tongue
FC - Foramen Cecum
SP - Soft Palate
E - Epiglottis
O - Opening of Larynx into Pharynx
VF - Level of Vocal Folds

Fig. 7b

DIAGRAM OF THE AIR PASSAGES OF NEANDERTHAL MAN

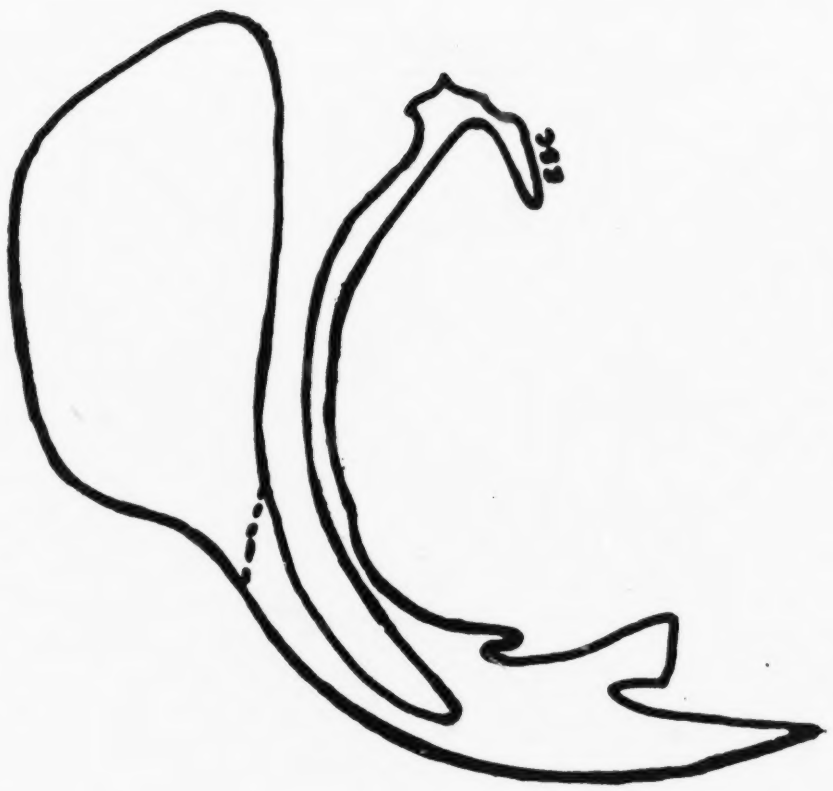
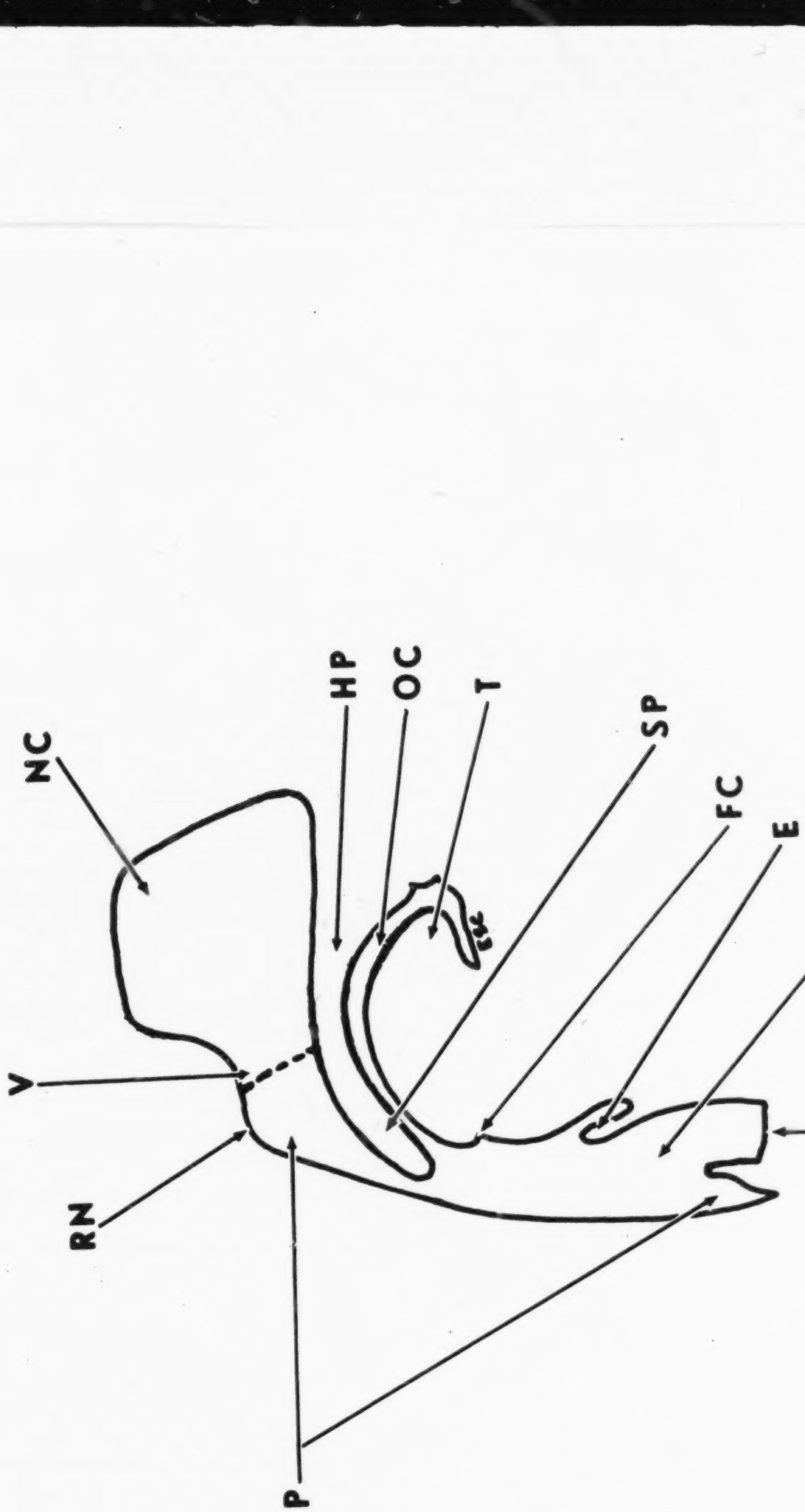


Fig. 7c

DIAGRAM OF THE AIR PASSAGES OF ADULT HUMAN



P - Pharynx
RN - Roof of Nasopharynx
V - Vomer Bone
NC - Nasal Cavity
HP - Hard Palate
OC - Oral Cavity
T - Tongue
FC - Foramen Cecum
SP - Soft Palate
E - Epiglottis
O - Opening of Larynx into Pharynx
VF - Level of Vocal Folds

Fig. 7d

SUPRALARYNGEAL VOCAL-TRACT CONSTRAINTS ON PHONETIC REPERTORIES

We have noted that human speech production involves a source of sound and a supralaryngeal vocal tract that acts as an acoustic "filter" or modulator. Man uses his articulators (the tongue, lips, mandible, velum, pharyngeal constrictors, etc.) to modify dynamically in time the resonant structure that the supralaryngeal vocal tract imposes on the acoustic sound pressure radiated at the speaker's lips and nares.

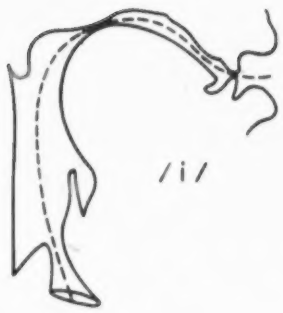
The phonetic inventory of a human language is therefore limited by (1) the number of acoustically distinct sound sources that man is capable of controlling during speech communication, and (2) the number of distinct resonant patterns available through positioning of the articulators and dynamic manipulation of the articulators. In most human languages, a phonetic analysis will reveal a phonemic inventory on the order of twenty to forty distinct sound types (Troubetzkoy, 1939; Jakobson et al., 1952). Most of the segment proliferations are achieved through the varied use of the articulators. For example, in English there are at least ten vowels that differ primarily in the articulatory configuration of the supralaryngeal vocal tract and concomitantly in the resonant, i.e., the formant structure of the acoustic output (Peterson and Barney, 1952).

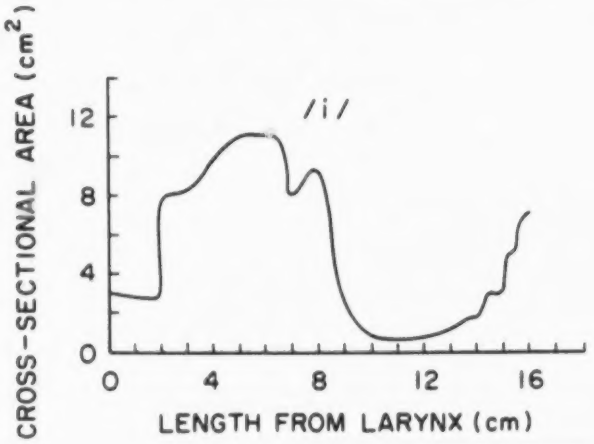
There is a direct relationship between the articulatory configuration of the supralaryngeal vocal tract and the formant structure (Fant, 1960). The relationship depends exclusively on the area function or cross-sectional area of the vocal tract as a function of the distance from the vocal cords to the lips. The availability of digital computers makes it possible to determine the range of formant frequency patterns that a supralaryngeal vocal tract can produce. If the supralaryngeal vocal-tract area function is systematically manipulated in accord with the muscular and anatomical constraints of the head and neck, a computer can be programmed to compute the formant frequencies that correspond to the total range of supralaryngeal vocal-tract variation (Henke, 1966). In other words, a computer-implemented model of a supralaryngeal vocal tract can be used to determine the possible contribution of the vocal tract to the phonetic repertoire. We can conveniently begin to determine whether a nonhuman supralaryngeal vocal tract can produce the range of sounds that occurs in human language by exploring its vowel-producing ability. Consonantal vocal-tract configurations can also be modelled. It is, however, reasonable to start with vowels since the production of consonants may also involve rapid, coordinated articulatory maneuvers and we can only speculate on the presence of this ability in fossil hominids.

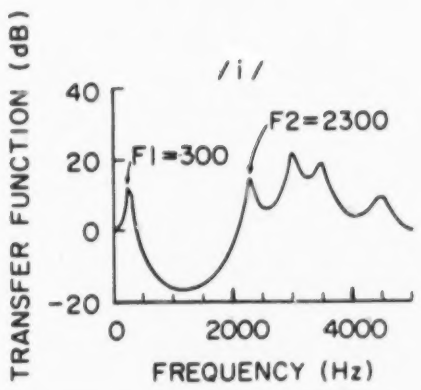
THE VOWEL TRIANGLE

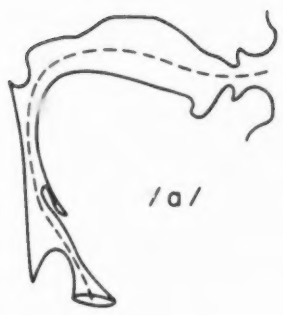
Articulatory and acoustic analyses have shown that the three vowels /i/, /a/, and /u/ are the limiting articulations of a vowel triangle that is language universal (Troubetzkoy, 1939). The body of the tongue is high and fronted to form a constricted oral cavity in /i/, whereas it is low to form a large oral cavity in /a/ and /u/. Figure 8 shows a midsaggital outline of the vocal tract for the vowels /i/, /a/, and /u/, as well as the cross-sectional areas of the vocal tract (Fant, 1960) and the frequency domain transfer functions for these vowels (Gold and Rabiner, 1968). The tongue body forms a large pharyngeal cavity in /i/ and /u/ and a constricted pharyngeal cavity in /a/. If

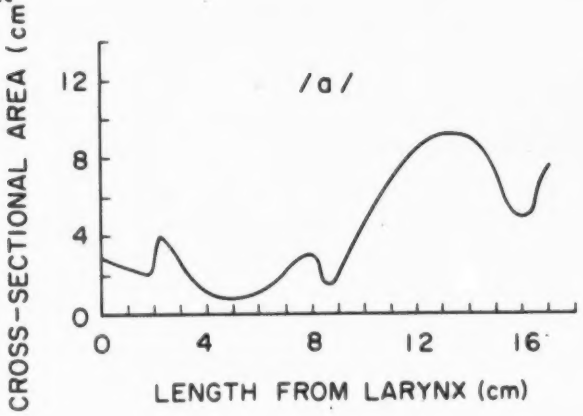
ILLUSTRATIONS OF APPROXIMATE MIDSAGGITAL SECTIONS, CROSS-SECTIONAL AREA FUNCTIONS, AND ACOUSTIC TRANSFER FUNCTIONS OF THE VOCAL TRACT FOR THE VOWELS /i/, /a/, AND /u/

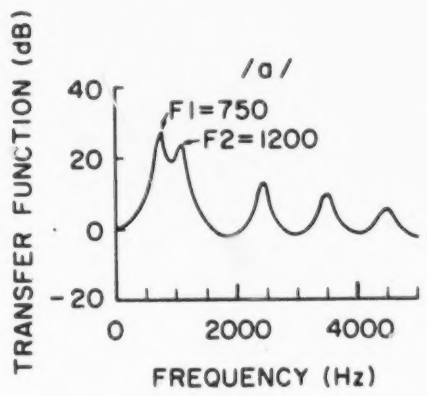


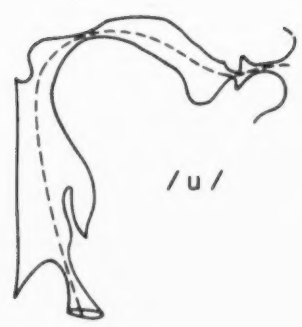


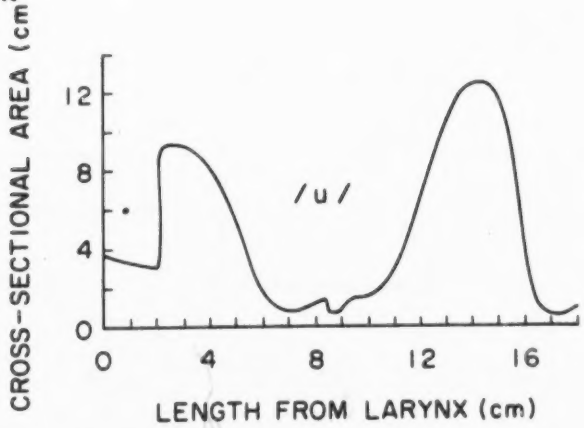


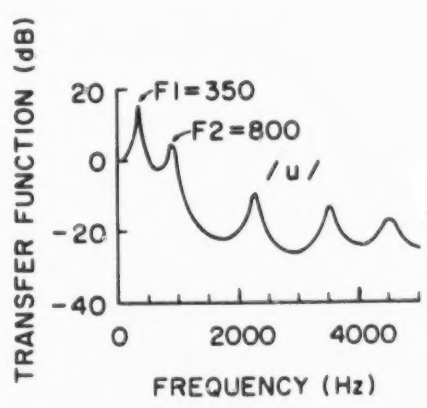












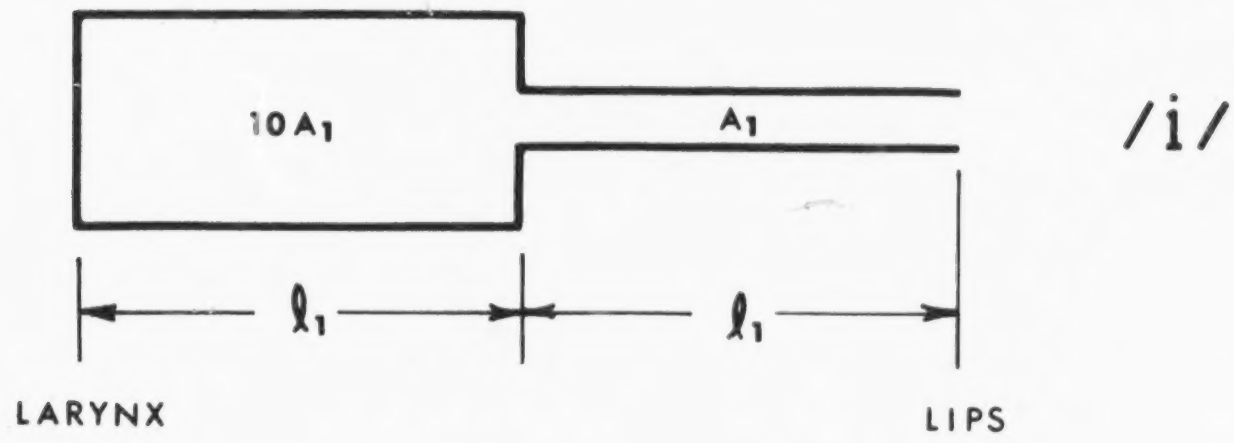
MIDSAGGITAL SECTION OF THE VOCAL TRACT

CROSS-SECTIONAL AREA FUNCTION OF THE VOCAL TRACT

MAGNITUDE OF THE VOCAL TRACT TRANSFER FUNCTION

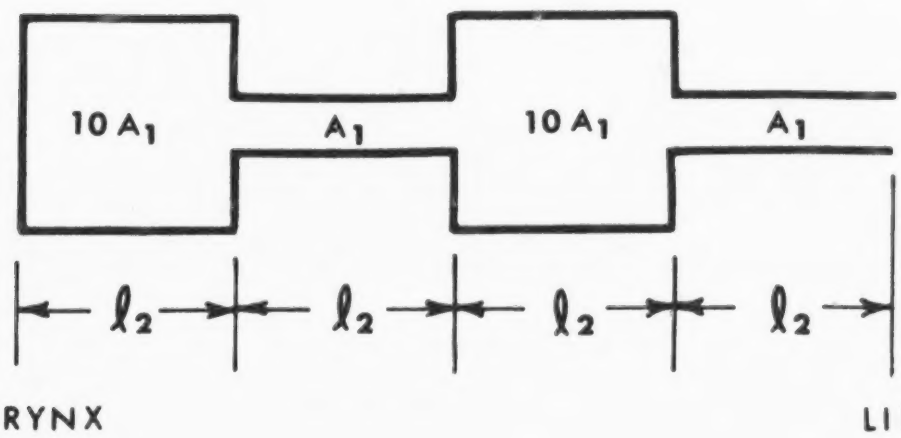
Fig. 8

STYLIZED SUPRALARYNGEAL VOCAL-TRACT AREA FUNCTIONS THAT CHARACTERIZE THE HUMAN VOWELS /a/, /i/, AND /u/



/i/

/a/



/u/

Fig. 9

the tongue body moves to form any greater constrictions, turbulent friction noise is generated at the vocal-tract constriction and the articulation produces a consonant, not a vowel. Other English vowels are produced by means of supralaryngeal vocal-tract configurations within the articulatory triangle[4] defined by /i/, /a/, and /u/.

The universality and special nature of /i/, /a/, and /u/ can be argued from theoretical grounds as well. Employing the simplified and idealized area functions shown in Figure 9, Stevens (1969) has shown that these articulatory configurations (1) are acoustically stable for small changes in articulation and therefore require less precision in articulatory control than similar adjacent articulations and (2) contain a prominent acoustic feature, i.e., two formants that are in close proximity to form a distinct energy concentration.

The vowels /a/, /i/, and /u/ have another unique property. They are the only vowels in which an acoustic pattern can be related to a unique vocal-tract area function (Lindblom and Sundberg, 1969; Stevens, 1969). Other vowels like /e/, /I/, /U/, etc., can be produced by means of several alternate area functions (Stevens and House, 1955). A human listener, when he hears a syllable that contains a token of /a/, /i/, or /u/, can calculate the size of the supralaryngeal vocal tract that was used to produce the syllable. The listener, in other words, can tell whether a speaker has a large or a small vocal tract. This is not possible for other vowels since a speaker with a small vocal tract can, for example, by increasing the degree of lip rounding, produce a token of /U/ that would be consistent with a larger vocal tract with less lip rounding. These uncertainties do not exist for /a/, /i/, and /u/ since the required discontinuities in the supralaryngeal vocal-tract area functions (Figure 8) produce acoustic patterns that are beyond the range of compensatory maneuvers. The degree of lip rounding for the /u/ in Figure 8 is, for example, so extreme that it is impossible to constrict the lip opening any more and still produce a vowel.[5] The vowels /a/, /i/ and /u/ are therefore different in kind from the remaining "central" vowels. These "vocal-tract size calibrating" properties of /a/, /i/, and /u/ have a crucial role in the perception of speech, and we will have more to say on this matter.

We can conclude from these considerations that the vowel space reserved for human language is delimited by the vowels /a/, /i/, and /u/. A study of the theoretical limitations on vowels produced by another related species can therefore proceed by determining the largest vowel triangle that its articulatory system is capable of generating.

THE VOWEL TRIANGLE IN CHIMPANZEE AND NEWBORN MAN AND NEANDERTHAL MAN

Some general observations are in order before detailed consideration of the vowel-producing capabilities of chimpanzee, human newborn, and Neanderthal man. The idealized area functions of Stevens (Figure 9) require a relatively

---

[4] It can be argued that / ɔ / forms a fourth position on a vowel "quadrangle," but this modification will not affect our arguments in any essential way.

[5] If the size of the constriction becomes too small, turbulent noise will be generated at the constriction and the sound will no longer be a vowel.

CHIMPANZEE SUPRALARYNGEAL VOCAL-TRACT AREA FUNCTIONS MODELLED ON COMPUTER

/i/ ●—●			/a/ ■– –■			/u/ ▲·····▲		
Formant	Freq.	Freq./1.7	Formant	Freq.	Freq./1.7	Formant	Freq.	Freq./1.7
1	610	360	1	1220	720	1	830	490
2	3400	2000	2	2550	1500	2	1800	1060
3	4420	2600	3	5070	2980	3	4080	2390

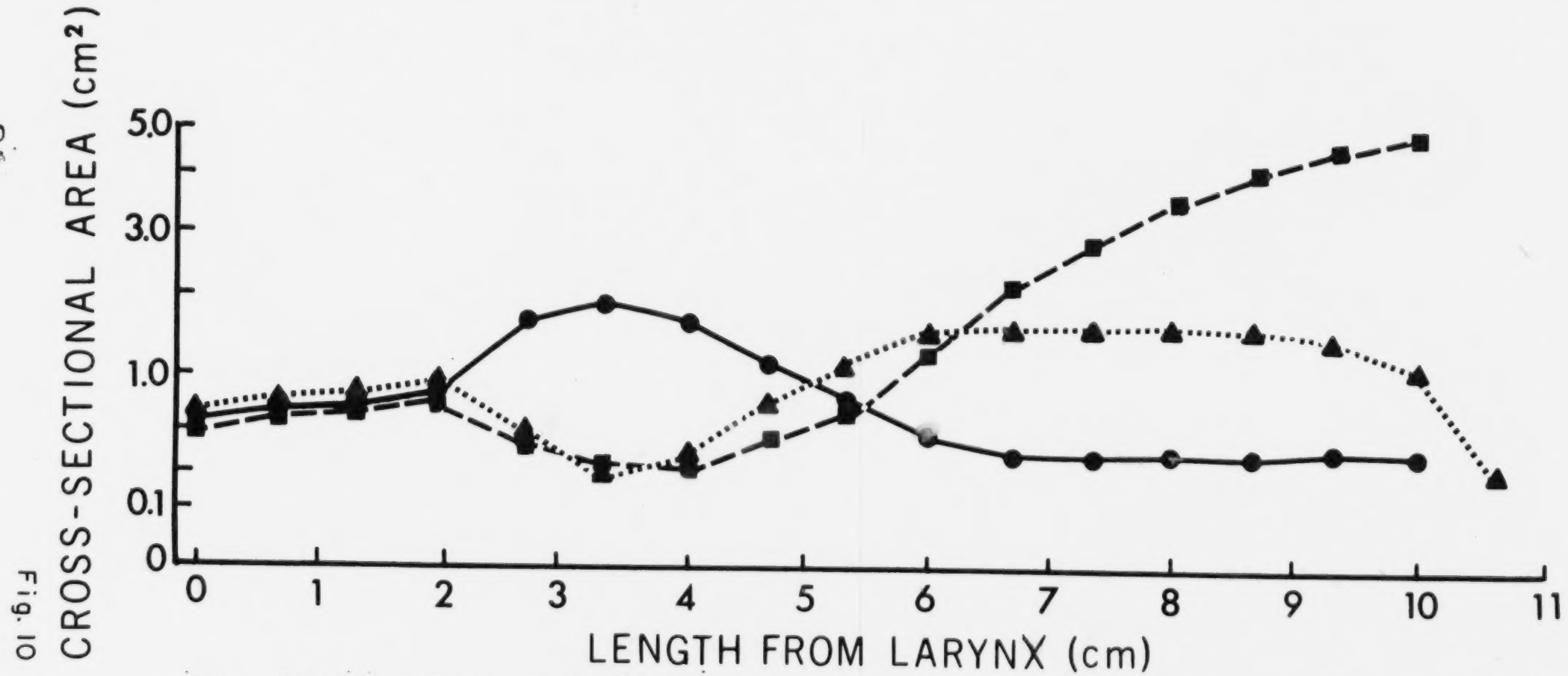


Fig. 10

Note: These functions were the "best" approximations that could be produced, given the anatomic limitations of the chimpanzee, to the human vowels /i/, /a/, and /u/. The formant frequencies calculated by the computer program for each vowel are tabulated and scaled to the average dimensions of the adult human vocal tract.

large ratio of the areas of the large and small section. In addition they require rather abrupt boundaries between sections. These configurations can be approximated in adult man at the junction of the pharyngeal and oral cavities where the styloglossus muscle can be effective in pulling the body of the tongue upwards and backwards in the direction of the nasopharynx (Sobotta-Figge, 1965; Perkell, 1969; Lieberman, 1970). The cross-sectional area of the oral and pharyngeal cavities can be independently manipulated in adult man (refer to Figure 8) while a midpoint constriction is maintained. The supralaryngeal vocal tract of adult man thus can, in effect, function as a "two-tube" system. The lack of a supralaryngeal pharyngeal region prevents chimpanzee, human newborn, and Neanderthal from employing these mechanisms. They can only attempt to distort the tongue body in the oral cavity to obtain changes in cross-sectional areas. The intrinsic musculature of the tongue severely limits the range of deformations that the tongue body can be expected to employ. Chimpanzee, human newborn, and Neanderthal man, in effect, have "single-tube" resonant systems.

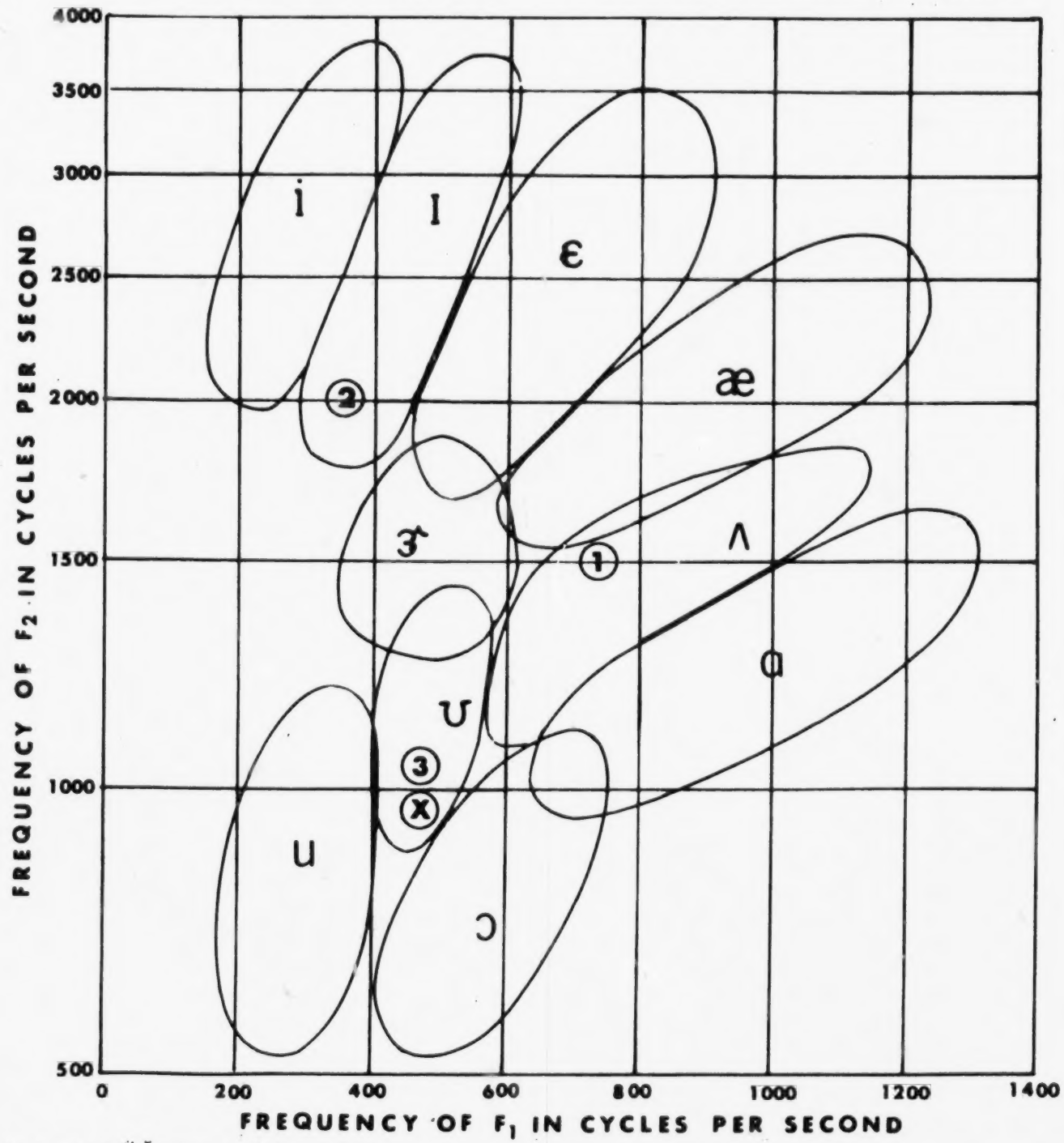
The chimpanzee and human newborn heads are both smaller than that of adult man. This imposes a further difficulty since it makes it difficult to form the large cavities that are found in the vowels of man. Therefore comparable cavity area ratios would require the use of smaller constrictions than adult man, but this would violate the requirement of nonturbulent flow in the constricted part of the vocal tract for vowels.

The Chimpanzee Vowel Triangle

The vowel /a/ could be articulated by a chimpanzee if he were to open his mandible sufficiently to obtain a flared area function. Taking into account the constraints mentioned above, an area function for a chimpanzee /a/ has been estimated and plotted in Figure 10. Formant frequencies corresponding to the area function have been computed by means of an algorithm described by Henke (1966) and are tabulated in the figure. The area of the vocal tract was specified at 0.5 cm intervals using this algorithm, which was implemented on a digital computer. When the two lowest formants are scaled down in frequency by a factor proportional to the ratio of a chimpanzee vocal-tract length of 10 cm to the mean vocal-tract length of 17 cm of adult man, then the chimpanzee formants can be compared directly with comparable data in adult man. This is done on a plot of first formant frequency versus second formant frequency in Figure 11 where the data point for this is denoted by the circled number "1." We see that the chimpanzee formant patterns for this vowel configuration do not fall within the range of /a/ data for man but rather lie inside the vowel triangle in the /ʌ/ region. The normative data for modern man with which the chimpanzee vowel is compared is derived from a sample of seventy-six adult men, adult women, and children (Peterson and Barney, 1952). The labelled loops enclose the data points that accounted for 90 percent of the samples in each vowel category. The children in the Peterson and Barney study were sufficiently old that they all had vocal tracts that conformed to that typical of adult morphology (Lieberman et al., 1968; Crelin and Lieberman, unpublished data).

The vowel /i/ could be best approximated by a chimpanzee by pulling the body of the tongue forward with the mandible lowered slightly. The cross-sectional area of the back cavity will not be large, but it may approach the area function estimated in Figure 10. This area function results in formant

PLOT OF FORMANT FREQUENCIES FOR CHIMPANZEE VOWELS OF FIGURE 9



Note: Data points (1), (2), and (3) are scaled to correspond to the size of the adult human vocal tract. Data point (X) represents an additional point for human newborn. The closed loops enclose 90 percent of the data points derived from a sample of seventy-six adult men, adult women, and children producing American-English vowels (Peterson and Barney, 1932). Note that the chimpanzee and newborn vocal tracts cannot produce the vowels /i/, /u/, and /a/.

Fig. 11

locations that are tabulated in Figure 10 and scaled and plotted in Figure 11 (data point "2"). The formants do not fall within the /i/ region in adult man but rather inside the vowel triangle in the /I/ region.

The vowel /u/ is virtually impossible for the chimpanzee to articulate. A large front cavity requires the mandible to be lowered because the simian shelf prevents the tongue body motion found in man. However, the required lip rounding is incompatible with a lowered mandible. An approximation to a chimpanzee /u/ area function is estimated in Figure 10. Again, the formant locations of this area function are computed, scaled, and plotted in Figure 11 (data point "3"). They indicate that the comparable English vowel is /U/ and not /u/.

The discussion of the vowel triangle has not considered the effects of the chimpanzee pharynx which acts as a relatively short side-branch resonator. The pharyngeal section may be essentially closed in a back vowel such as /a/, but it probably plays an important role in /i/. The presence of a side-branch resonator has the effect of modifying formant locations and also the effect of introducing antiresonances into the vocal-tract transfer function. We estimate that the lowest frequency antiresonance for /i/ of a slightly flared 6 cm pharyngeal section is about 2000 Hz.[6]

Newborn Human

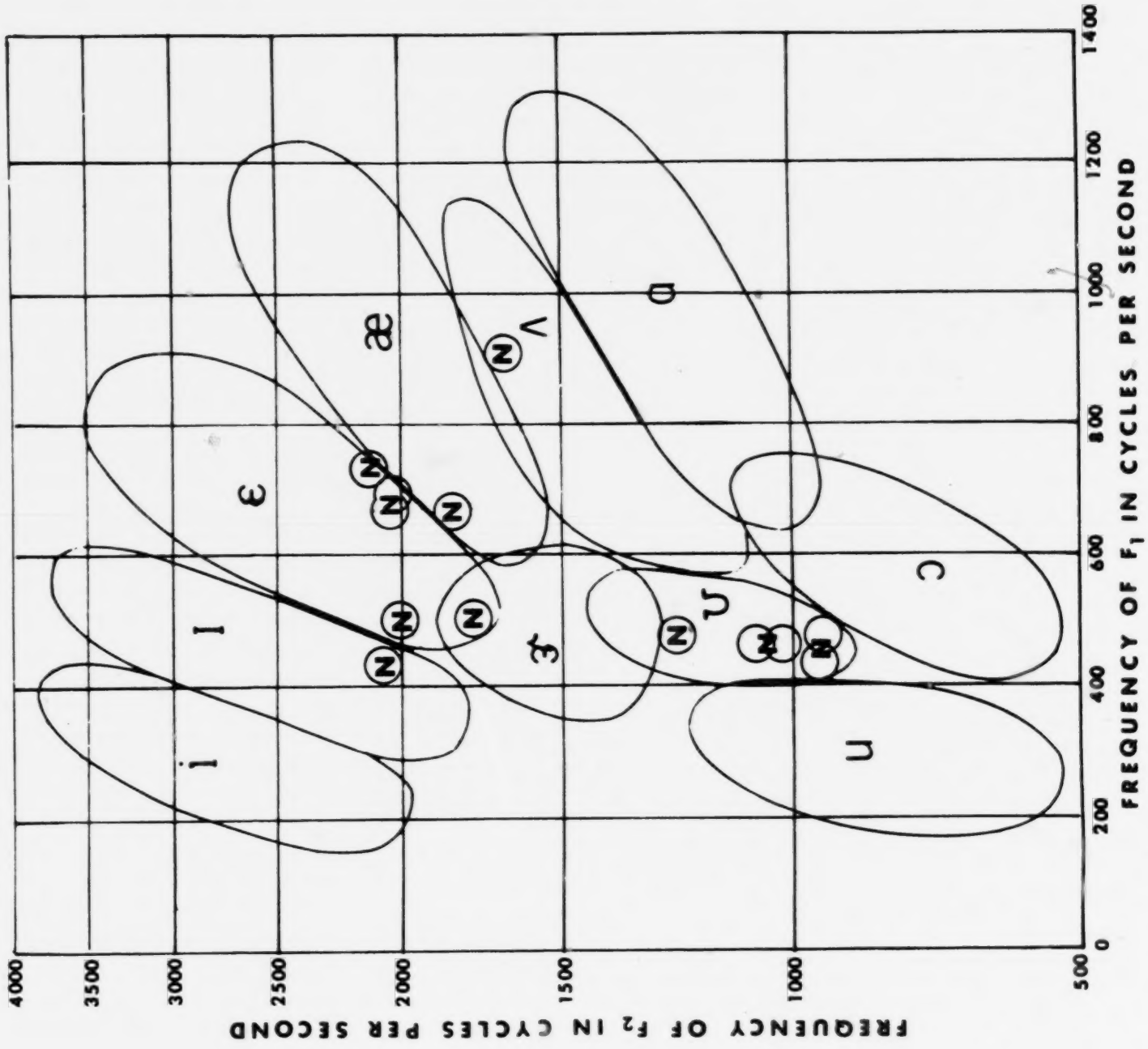
The supralaryngeal vocal tract of the human newborn does not differ substantially from the chimpanzee's (Figures 1 through 3). The absence of a simian shelf in the mandible, however, allows the formation of a larger front cavity in the production of vowels that approximate the adult human /u/. In Figure 11 the formant locations of this area function, which resemble that of Figure 10 for the chimpanzee /u/ approximation with a larger front cavity, are computed, scaled, and plotted as data point "X." The resulting vowel sound is comparable to the English vowel /U/ not /u/, but it is a closer acoustic approximation to /u/. The acoustic output of the newborn vocal tract does not otherwise differ substantially from the chimpanzee vocal tract. Perceptual and acoustic studies of the vocalizations of human newborn (Irwin, 1957; Lieberman et al., 1968) show that all, and only, the vowels that can be produced are indeed produced.

Neanderthal Man

The vowel-producing abilities of the reconstructed supralaryngeal vocal tract of the La Chapelle-aux-Saints Neanderthal fossil are presented in Figure 12. The formant frequencies of the Neanderthal supralaryngeal vocal-tract configurations that best approximated the human vowels /a/, /i/, and /u/ were computed, scaled, and plotted with respect to adult modern man (Lieberman and Crelin, 1971). Note that the Neanderthal vowels, which are each labelled "N," do not fall in the human ranges for /a/, /i/, or /u/. The Neanderthal vocal tract was given the benefit of all possible doubts in the computer modelling. The maximum range of laryngeal cavity variation typical of modern man (Fant, 1960) was, for example, used in a manner that would enhance the phonetic

[6]This may have a perceptual effect similar to that of nasality as transfer function zeros appear in adult human speech in nasalized vowels.

PLOT OF FIRST AND SECOND FORMANT FREQUENCIES FOR "EXTREME" VOWELS,
DATA POINTS (N), OF RECONSTRUCTED NEANDERTHAL VOCAL TRACT



(Lieberman and Crelin, 1971)

Fig. 12

ability of the Neanderthal vocal tract. Articulatory maneuvers that would be somewhat acrobatic in modern man were also used to enhance Neanderthal phonetic ability. Our computer modelling was guided by the results of X-ray motion pictures of speech production, vocalization, swallowing, and respiration in adult man (Perkell, 1969; Haskins Laboratories, 1962) and in newborn (Truby et al., 1965). This knowledge, plus the known comparative anatomy of the living primates, allowed a fairly "conservative" simulation of the vowel-producing ability of this fossil specimen who is typical of the range of "classic" Neanderthal man.[7] We perhaps allowed a greater vowel-producing range for Neanderthal man since we consistently generated area functions that were more human-like than ape-like whenever we were in doubt. Despite these compensations, the Neanderthal vocal tract cannot produce /a/, /i/, or /u/. The absence of these vowels from the vowel systems of chimpanzee, newborn human, and Neanderthal man in Figures 11 and 12 thus is an indirect way of showing that the vocal tracts of these creatures cannot form the abrupt area functions that are necessary for these vowels. Our modelling of the newborn vocal tract served as a control procedure since we were able to produce the vowels that newborn humans actually produce. We produced, however, a greater vowel range than has been observed in the acoustic analysis of chimpanzee vocalizations (Lieberman, 1968). We will return to this point later in our discussion since it may reflect the absence of required neural mechanisms in the nonhuman primates.

SPEECH PRODUCTION AND SPEECH PERCEPTION

Supralaryngeal vocal-tract area functions that approximated typical consonantal configurations for adult man (Fant, 1960; Perkell, 1969) were also modelled on the digital computer (Lieberman and Crelin, 1971). Chimpanzee, newborn human, and Neanderthal man all appeared to have anatomical mechanisms that would allow the production of both labial and dental consonants like /b/, /p/, /t/, /s/, etc., if other muscular and neural factors were present.

It is obvious that some of these factors are not present in newborn human since neither labial nor dental consonants occur in the utterances of newborn infants (Irwin, 1957). It is possible that the nonoccurrence of these consonants is a consequence of a general inability to produce rapid articulatory maneuvers. The situation is more complex in chimpanzee where a discrepancy again exists between the constraints that the supralaryngeal vocal tract imposes on the phonetic repertoire and actual performance. Chimpanzees do not appear to produce dental consonants, although they have the anatomical "machinery" that would permit them to do so. Observations of captive chimpanzees have not, for example, revealed patterns of vocal communication that utilize contrasts between labial and dental consonants (Lieberman, 1968). It is unlikely that the failure to observe dental consonants in chimpanzee

---

[7] We have noted (Lieberman and Crelin, 1971) that a number of fossils, which differ slightly in other ways, all have a "flattened-out" skull base and other anatomical features that indicate the absence of a supralaryngeal vocal tract like adult modern man's. There is, in other words, a class of "Neanderthaloid" fossils that lacks the ability to produce the full range of human speech.

vocalizations is due to a limited data sample since attempts to train chimpanzees to mimic human speech have not succeeded in teaching them to produce dental consonants. At least one chimpanzee has been taught to produce labial consonants like /p/ and /m/ (Hayes, 1952) so the absence of dental consonants cannot be ascribed to a general inability to produce rapid articulatory maneuvers.

Our computer modelling of the chimpanzee vocal tract shows that these animals have the anatomic ability that would allow them to produce a number of vowels that in human speech are "phonemic" elements, i.e., sound contrasts that convey linguistically meaningful information. Chimpanzees, however, do not appear to make use of these vowel possibilities. Instead, they appear to make maximum use of the "neutral," uniform cross-section supralaryngeal vocal-tract shape (Jakobson et al., 1952; Lieberman, 1968) with source variations. Chimpanzees, for example, will make calls that are different insofar as the glottal excitation is weak, breathy, has a high fundamental frequency, etc.[8]

The absence of sounds that are anatomically possible may perhaps reflect perceptual limitations. In other words, chimpanzees may not use dental consonants in contrast with labial consonants because they cannot perceptually differentiate these sounds. Differences in vowel quality as between /I/ and /e/, for example, may also be irrelevant for chimpanzees. The absence of the vowels /a/, /i/, and /u/ from the chimpanzee's phonetic abilities is consistent with this hypothesis which has wider implications concerning the general phonetic and linguistic abilities of the living nonhuman primates and hominid fossils like Neanderthal man.

SPEECH AND LANGUAGE

Linguists have, as we noted earlier, tended to ignore the phonetic level of language and speech production. The prevailing assumption is that the interesting action is at the syntactic and semantic levels and that just about any sequence of arbitrary sounds would do for the transfer of linguistic information. Some linguists might, for example, point out that even simple binary codes, such as Morse code, can be used to transmit linguistic information. Neanderthal man, in this view, therefore would need only one sound contrast to communicate. After all, modern man can communicate by this means: why not Neanderthal man? The answer to this question is quite simple. Human speech is a special mode of communication that allows modern man to communicate at least ten times faster than any other known method. Sounds other than speech cannot be made to convey language well.[9] That knowledge comes from 55 years of trying to make nonspeech sounds for use in reading machines for the blind, that is, devices that scan the print and convert it into meaningful sounds. In spite of the most diligent efforts in connection with the development

---

[8] Meaningful chimpanzee calls can be "seen" in context in the recent sound motion pictures taken by P. Marler at the Gombe Stream Reserve chimpanzee project of J. Goodall (1965).

[9] I am essentially paraphrasing the discussion presented by A.M. Liberman (1970) with regard to the linguistic status of human speech and the process of speech encoding. Liberman's logic is clear, correct, and succinct.

of these machines, no nonspeech acoustic alphabet has yet been contrived that can be made to work more than one-tenth as well as speech (Liberman et al., 1967). Nor has any better degree of success attended efforts towards the use of visual displays in the development of "hearing" machines for the deaf (Koenig et al., 1946).

The problem is quite clear when one considers the rate at which information is transferred in human speech. Human listeners can perceive as many as twenty-five to thirty phonetic segments per second in normal speech. This information rate far exceeds the resolving power of the human auditory system. It is, for example, impossible even to count simple pulses at rates of twenty pulses per second. The pulses simply merge into a continuous tone. Communication by means of Morse code would be possible, but it would be very slow. Human speech achieves its high information rate by means of an "encoding" process that in structured in terms of the anatomic and articulatory constraints of speech production. The presence of vowels like /a/, /i/, and /u/ appears to be one of the anatomic factors that makes this encoding process possible.

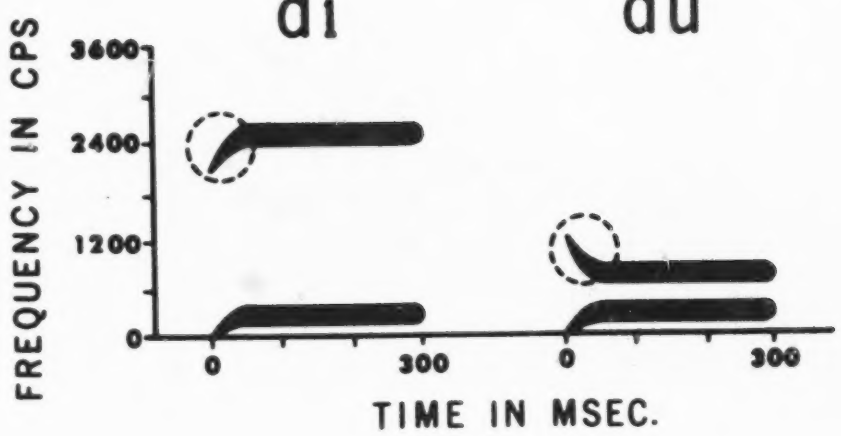
SPEECH ENCODING AND THE "MOTOR THEORY" OF SPEECH PERCEPTION

In human speech a high rate of information transfer is achieved by "encoding" phonetic segments into syllable-sized units. The phonetic representation of a syllable like /du/ essentially states that two independent elements are being transmitted. The syllable /du/ can be segmented at the phonetic level into two segments, /d/ and /u/, which can independently combine with other phonetic segments to form syllables like /di/ or /gu/. Phonetic segments like /d/, /g/, /u/, and /i/ are also independent at the articulatory level insofar as these phonetic elements can each be specified in terms of an articulatory configuration. The phonetic element /u/ thus involves a particular vocal-tract configuration which approximates that in Figure 8. The phonetic element /d/ likewise involves a particular vocal-tract configuration in which the tongue blade momentarily occludes the oral cavity. It is possible to effect a segmentation of the syllable /du/ at the articulatory level. If an X-ray motion picture of a speaker producing the syllable /du/ were viewed it would, for example, be possible to see the articulatory gesture that produces the /d/ in the syllable /du/. It is not, however, possible to segment the acoustic correlates of /d/ from the speech signal.

In Figure 13, we have reproduced two simplified spectrographic patterns that will, when converted to sound, produce approximations to the syllables /di/ and /du/ (Liberman, 1970).[10] The dark bands on these patterns represent the first and second formant frequencies of the supralaryngeal vocal tract as functions of time. Note that the formants rapidly move through a range of frequencies at the left of each pattern. These rapid movements, which occur in about 50 msec, are called formant transitions. The transition in the second formant, which is encircled, conveys the acoustic information that human listeners interpret as a token of a /d/ in the syllables /di/ and

[10] It can be argued that the primary acoustic cue to the identity of /d/ is a brief high-frequency burst of fricative noise. However adult listeners will respond correctly to the acoustic signals defined in Figure 13 even though this cue is missing.

SIMPLIFIED SPECTROGRAPHIC PATTERNS SUFFICIENT TO PRODUCE
THE SYLLABLES /di/ AND /du/



Note: The circles enclose the second formant frequency transitions.
(After Liberman, 1970).

Fig. 13

/du/. It is, however, impossible to isolate the acoustic pattern of /d/ in these syllables. If tape recordings of these two syllables are "sliced" with the electronic equivalent of a pair of scissors (Lieberman, 1963), it is impossible to find a segment that contains only /d/. There is no way to cut the tape so as to obtain a piece that will produce /d/ without also producing the next vowel or some reduced approximation to it.

Note that the encircled transitions are different for the two syllables. If these encircled transitions are isolated, listeners report that they hear either an upgoing or a falling frequency modulation. In context, with the acoustic correlates of the entire syllable, these transitions cause listeners to hear an "identical" sounding /d/ in both syllables. How does a human listener effect this perceptual response?

We have noted the formant frequency patterns of speech reflect the resonances of the supralaryngeal vocal tract. The formant patterns that define the syllable /di/ in Figure 13 thus reflect the changing resonant pattern of the supralaryngeal vocal tract as the speaker moves his articulators from the occlusion of the tongue tip against the palate that is involved in the production of /d/ to the vocal-tract configuration of the /i/. A different acoustic pattern defines the /d/ in the syllable /du/. The resonances of the vocal tract are similar as the speaker forms the initial occlusion of the /d/ in both syllables; however, the resonances of the vocal tract are quite different for the final configurations of the vocal tract for /i/ and /u/. The formant patterns that convey the /d/ in both syllables are thus quite different since they involve transitions from the same starting point to different end points. Human listeners "hear" an identical initial /d/ segment in both of these signals because they "decode" the acoustic pattern in terms of the articulatory gestures and the anatomical apparatus that is involved in the production of speech. The listener in this process, which has been termed the "motor theory of speech perception" (Liberman et al., 1967), operates in terms of the acoustic pattern of the entire syllable. The acoustic cues for the individual "phonetic segments" are fused into a syllabic pattern. The high rate of information transfer of human speech is thus due to the transmission of acoustic information in syllable-sized units. The phonetic elements of each syllable are "encoded" into a single acoustic pattern which is then "decoded" by the listener to yield the phonetic representation.

In order for the process of "motor theory perception" to work, the listener must be able to determine the absolute size of the speaker's vocal tract. Similar articulatory gestures will have different acoustic correlates in different sized vocal tracts. The frequency of the first formant of /a/, for example, varies from 730 to 1030 Hz in the data of Peterson and Barney (1952) for adult men and children. The frequencies of the resonances that occur for various consonants likewise are a function of the size of the speakers' vocal tract. The resonant pattern that is the correlate of the consonant /g/ for a speaker with a large vocal tract may overlap with the resonant pattern of the consonant /d/ for a speaker with a small vocal tract (Rand, ms.). The listener therefore must be able to deduce the size of the speaker's vocal tract before he can assign an acoustic signal to the correct consonantal or vocalic class.

There are a number of ways in which a human listener can infer the size of a speaker's supralaryngeal vocal tract. He can, for example, note the fundamental frequency of phonation. Children, who have smaller vocal tracts, usually have higher fundamental frequencies than adult men or adult women. Adult men, however, have disproportionately lower fundamental frequencies than adult women (Peterson and Barney, 1952), so fundamental frequency is not an infallible cue to vocal-tract size. Perceptual experiments (Ladefoged and Broadbent, 1957) have shown that human listeners can make use of the formant frequency range of a short passage of speech to arrive at an estimate of the size of a speaker's vocal tract. Recent experiments, however, show that human listeners do not have to defer their "motor theory" decoding of speech until they hear a two or three second interval of speech. Instead, they use the vocalic information encoded in a syllable to decode the syllable (Darwin, 1971; Rand, ms). This may appear to be paradoxical, but it is not. The listener makes use of the formant frequencies and fundamental frequency of the syllable's vowel to assess the size of the vocal tract that produced the syllable. We have noted throughout this paper that the vowels /a/, /i/, and /u/ have a unique acoustical property. The formant frequency pattern for these vowels can always be related to a unique vocal-tract size and shape.[11] A listener, when he hears one of these vowels, can thus instantly determine the size of the speaker's vocal tract. The vowels /a/, /i/, and /u/ (and the glides /y/ and /w/) thereby serve as primary acoustic calibration signals in human speech.

The anatomical impossibility for the chimpanzee to produce these vowels is thus consistent with the absence of meaningful changes in vowel quality in the vocal communications of these animals. Chimpanzees probably cannot perceive these differences in vowel quality because they cannot "decode" specific vowels and consonants in terms of the articulatory gestures that speakers use to produce these signals. A chimpanzee on hearing a particular formant frequency pattern would, for example, not be able to tell whether it was produced by a large chimpanzee who was using an /I/-like vocal-tract configuration or a smaller chimpanzee who was using an /e/-like vocal-tract configuration.[12] Chimpanzees simply may not have the neural mechanism that is

---

[11] The exact size and shape of the vocal tract can be theoretically calculated from the formant frequency pattern of these vowels if all of the theoretically infinite number of formant frequencies are known. If one, however, assumes that the formant structure of an unknown vowel is similar to /i/, /u/, or /a/ and is produced by a cavity shape shown in Figure 9, then the two lowest formants give a good estimate of vocal-tract length and size. The "quantal" nature of the speech signal discussed by Stevens (1969) makes an "exact" knowledge of vocal-tract size unnecessary for speech decoding.

[12] The Ladefoged and Broadbent (1957) vowel perception study is very pertinent in this regard since it shows that human listeners also cannot tell whether the acoustic signal that is a token of a "central" vowel is an /U/, an /I/, or an /ɚ/ in the absence of information that tells them the size of the speaker's vocal tract. The listeners in this experiment said that the same acoustic signal "was" the word bit, bat, or but when prior acoustic context led them to believe that the speaker had a large, medium, or small supralaryngeal vocal tract.

used in modern man to decode speech signals in terms of the underlying articulatory maneuvers. The absence of a human-like pharyngeal region in chimpanzee is thus quite reasonable. The only function that the human supralaryngeal vocal tract is better adapted to is speech production, in particular the production of vowels like /a/, /i/, and /u/. The adult human supralaryngeal vocal tract is otherwise less well adapted for the primary vegetative functions of swallowing and respiration (Negus, 1949). It is quite easy for food to be caught in the adult human pharynx and block the entrance to the larynx with fatal consequences, whereas the high position of the laryngeal opening in chimpanzees and other nonhuman primates would allow them to breathe with food lodged in their pharynx. The efficiency of the respiratory apparatus is reduced considerably in adult human because the angulation of the airway (Figures 6 and 7), resulting from the low position of the larynx, appreciably lessens the volume of air which could pass through a straight tube of equal cross-section. The high position of the larynx in newborn human, chimpanzee, and Neanderthal man is efficient for respiration. As Kirchner (1970) notes, "the larynx of the newborn infant is, from the standpoint of position, a more efficient respiratory organ than its adult counterpart."

This suggests that the evolution of the human vocal tract which allows vowels like /a/, /i/, and /u/ to be produced, and the universal occurrence of these vowels in human languages, reflect a parallel development of the neural and anatomic abilities that are necessary for language. This parallel development would be consistent with the evolution of other human abilities. The ability to use tools depends, for example, both on upright posture and an opposable thumb and on neural ability. As Darwin (1859, p. 194) noted, the theory of evolution through natural selection "can act only by taking advantage of slight successive variations; she can never take a leap, but must advance by the shortest and slowest steps." We can think of a process in which mutations that enhanced vocal communication were retained. The presence of enhanced mental ability would enhance the probability of the retention through natural selection of an anatomical mutation that enhanced the phonetic repertoire and the rate of communication. The presence of enhanced anatomical phonetic ability would, in turn, increase the probability of the retention of mutations that enhanced the neural abilities that are involved in speech encoding, decoding, syntax, etc. Positive feedback would, no doubt, result from this "circular" process. We would expect to find fossil forms like the La Chapelle-aux-Saints Neanderthal man who lacked a well-developed vocal mechanism but who undoubtedly must have had a "language." The remains of Neanderthal culture all point to the presence of linguistic ability.[13]

Note that a chimpanzee's response to simple human verbal requests does not demonstrate that the chimpanzee can "decode" human speech. The chimpanzee may be responding to acoustic factors that are not primary linguistic units, e.g., the prosodic features that relate to the emotionally determined "tone" of the speaker's voice. Psychoacoustic experiments designed to show whether nonhuman primates can "decode" speech have so far yielded negative results. It is indeed almost impossible to get nonhuman primates to respond to auditory signals wherever they readily respond to visual signals. (Kellogg, 1968; Hewes, 1971).

[13] Note that the prior existence of a form of language is necessary condition for the retention, through the process of natural selection, of mutations like the human pharyngeal region that enhance the rate of communication but are detrimental with regard to deglutition and respiration.

Neanderthal man lacked the vocal tract that is necessary to produce the human "vocal-tract size-calibrating" vowels /a/, /i/, and /u/. This suggests that the speech of Neanderthal man did not make use of syllabic encoding. While communication is obviously possible without syllabic encoding, studies of alternate methods of communication in modern man show, as we noted before, that the rate at which information can be transferred is about one-tenth that of normal human speech. The principle of encoding extends throughout the grammar of human languages. The process wherein a deep phrase marker with many elementary S's is collapsed into a derived surface structure may be viewed as an encoding process that is similar to the encoding that occurs between the phonetic level and speech (Liberman, 1970). A transformational grammar (Chomsky, 1957, 1965) may be viewed as a mechanism that encodes strings of semantic units into a surface structure. The derived surface string can be readily transmitted by a speaker and perceived and stored in short-time-span memory by a listener. There is no other reason why adult humans do not speak in short sentences like I saw the boy. The boy is fat. The boy fell down. instead of the "encoded" sentence I saw the fat boy who fell down. The "encoded" sentence can be transmitted more rapidly and it transmits the unitary reference of the single boy within the single breath-group (Lieberman, 1967). It thus is likely that Neanderthal man's linguistic abilities were at best suited to communication at slow rates and at worst markedly inferior at the syntactic and semantic levels to modern man's linguistic ability. Neanderthal man's language is an intermediate stage in the evolution of language. It may well have employed gestural communication as well as vocal signals (Hewes, 1971).

Human linguistic ability thus must be viewed as the result of a long evolutionary process that involved changes in anatomical structure through a process of mutation and natural selection which enhanced speech communication.[14] Modern man's linguistic ability is necessarily tied to his phonetic ability. Rapid information transfer through the medium of human speech must be viewed as a central property of human linguistic ability. It makes human language and human thought possible.

REFERENCES

Benda, C.E. (1969) Down's Syndrome, Mongolism and Its Management. Grune and Stratton, New York.

Boule, M. (1911-1913) "L'Homme fossile de la Chapelle-aux-Saints," Annales de Paleontologie, 6.109; 7.21, 85; 8.1.

Chiba, T. and M. Kajiyama (1958) The Vowel, Its Nature and Structure. Phonetic Society of Japan, Tokyo.

Chomsky, N. (1957) Syntactic Structure. Mouton, The Hague.

Chomsky, N. (1965) Aspects of the Theory of Syntax. M.I.T. Press, Cambridge Mass.

---

[14] We, therefore, see the evolution of language as a process that is ultimately based on mechanisms that exist in the more "primative" fossil hominids. It is probable that the living apes, as well as other animals, still have similar mechanisms. Studies of animal communication therefore are relevant to the study of human linguistic ability. We obviously do not agree with the theory that bases modern man's linguistic abilities on "unique" mechanisms that require "discontinuities" in evolution (Lenneburg, 1967).

Crelin, E.S. (1969) Anatomy of the Newborn; an Atlas. Lea and Febiger, Philadelphia.
Darwin, C. (1859) On the Origin of Species. (Facsimile edition) Atheneum, New York.
Darwin, C. (1971) "Ear Differences in the Recall of Fricatives and Vowels," Quarterly J. Exp. Psychol. 23.
Fant, C.G.M. (1960) Acoustic Theory of Speech Production. Mouton, The Hague.
Gold, B. and L.R. Rabiner (1968) "Analysis of Digital and Analog Formant Synthesizers," IEEE-Trans. Audio Electroacoustics, AU-16.81-94.
Goodall, J. (1965) "Chimpanzees of the Gombe Stream Reserve," in Primate Behavior, I. DeVore, ed. Holt, Rinehart and Winston, New York.
Greenewalt, C.A. (1967) Bird Song: Acoustics and Physiology. Smithsonian, Washington, D.C.
Haskins Laboratories (1962) X-Ray Motion Pictures of Speech. Haskins Laboratories, 305 E. 43 St., New York City.
Hayes, C. (1952) The Ape in Our House. Harper and Brothers, New York.
Henke, W.L. (1966) Dynamic Articulatory Model of Speech Production Using Computer Simulation. Doctoral dissertation, MIT (appendix B).
Hewes, G.W. (1971) Language Origins: A Bibliography. Dept. of Anthropology, Univ. of Colorado, Boulder.
Irwin, O.C. (1957) "Speech Development in Childhood," in Manual of Phonetics, L. Kaiser, ed. North-Holland, Amsterdam.
Jakobson, R., C.G.M. Fant, and M. Halle. (1952) Preliminaries to Speech Analysis. MIT Press, Cambridge.
Kellogg, W.N. (1968) "Communication and Language in the Home-Raised Chimpanzee," Science 162.423-427.
Kirchner, J.A. (1970) Pressman and Kelemen's Physiology of the Larynx, revised edition, Amer. Acad. Opthal. and Otolaryn., Rochester, Minn.
Koenig, W., H.K. Dunn, and L.Y. Lacy (1946) "The Sound Spectrograph," J. Acoustical Society America 17.19-49.
Ladefoged, P. and D.E. Broadbent (1957) "Information Conveyed by Vowels," J. Acoust. Soc. Am. 29.98-104.
Laughlin, W.S. (1963) "Eskimos and Aleuts: Their Origins and Evolution," Science 142.633-645.
LaMettrie, J.O. (1747) De L'Homme-machine. A. Vartanian, ed., Princeton Univ. Press, Princeton, N.J. (1960 critical edition).
Lenneburg, E.H. (1967) Biological Foundations of Language. Wiley, New York.
Liberman, A.M. (1970) "The Grammars of Speech and Language," Cognitive Psychology 1.301-323.
Liberman, A.M., D.P. Shankweiler, and M. Studdert-Kennedy (1967) "Perception of the Speech Code," Psychol. Rev. 74.431-461.
Lieberman, P. (1963) "Some Effects of Semantic and Grammatical Context on the Production and Perception of Speech," Language and Speech 6.172-187.
Lieberman, P. (1967) Intonation, Perception and Language. MIT Press, Cambridge.
Lieberman, P. (1968) "Primate Vocalizations and Human Linguistic Ability," J. Acoust. Soc. Am. 44.1574-1584.
Lieberman, P. (1969) "On the Acoustic Analysis of Primate Vocalizations," Behav. Res. Meth. and Instru. 1.169-174.
Lieberman, P. (1970) "Review of Perkell's (1968) Physiology of Speech Production," Language Sciences 13.25-28.
Lieberman, P., K.S. Harris, P. Wolff, and D.H. Russell (1968) "Newborn Infant Cry and Nonhuman Primate Vocalizations," Status Report 17/18, Haskins Laboratories, New York City.

Lieberman, P., D.H. Klatt, and W.A. Wilson (1969) "Vocal Tract Limitations of the Vocal Repertories of Rhesus Monkey and Other Non-human Primates," Science 164.1185-1187.

Lieberman, P. and E.S. Crelin (1971) "On the Speech of Neanderthal Man," Linguistic Inquiry 2, No. 2.

Lindblom, B. and J. Sundberg (1969) "A Quantitative Model of Vowel Production and the Distinctive Features of Swedish Vowels," Speech Transmission Laboratory Report 1, Royal Institute of Technology, Stockholm, Sweden.

Negus, V.E. (1949) The Comparative Anatomy and Physiology of the Larynx. Hafner, New York.

Patte, E. (1955) Les Neanderthaliens, Anatomie, Physiologie, Comparaisons. Masson et cie., Paris.

Perkell, J.S. (1969) Physiology of Speech Production: Results and Implications of a Quantitative Cineradiographic Study. MIT Press, Cambridge.

Peterson, G.E. and H.L. Barney (1952) "Control Methods Used in a Study of the Vowels," J. Acoust. Soc. Am. 24.175-184.

Rand, T.C. (ms) "Vocal Tract Size Normalization in the Perception of Stop Consonants."

Schultz, A.H. (1968) "The Recent Hominoid Primates," in Perspectives on Human Evolution, S.L. Washburn and Phyllis C. Jay. Holt, Rinehart and Winston, New York.

Simpson, G.G. (1966) "The Biological Nature of Man," Science 152.472-478.

Sobotta-Figge (1965) J. Sobotta and F.H.J. Figge, Atlas of Human Anatomy, Vol. II. Hafner, New York.

Stevens, K.N. (1969) "The Quantal Nature of Speech: Evidence from Articulatory-Acoustic Data," in Human Communication: A Unified View, E.E. David, Jr. and P.B. Denes, eds. McGraw-Hill, New York.

Stevens, K.N. and A.S. House (1955) "Development of a Quantitative Description of Vowel Articulation," J. Acoust. Soc. Am. 27.484-493.

Troubetzkoy, N.S. (1939) Principes de phonologie. Trans. 1949, J. Cantineau. Klincksieck, Paris.

Truby, H.M., J.F. Bosman, and J. Lind (1965) Newborn Infant Cry. Almqvist and Wiksells, Uppsala.

Virchow, R. (1872) "Untersuchung des Neanderthal--Schadels" Z. Ethan. 9.152-165.

A Voice for the Laboratory Computer*

F.S. Cooper, T.C. Rand, R.S. Music, and I.G. Mattingly+
Haskins Laboratories, New Haven

It seems clear, even from the titles for this session, that voice answer-back from computers can be used in diverse ways. But responding by voice is by no means all that the ability to speak will enable a computer to do. One of these other things is to read aloud for the blind. This possibility--even most of the basic methods for realizing it--was clearly foreseen well before the technology was ready (Cooper, 1963). It is a challenging problem for several reasons: first, because it is so completely open-ended linguistically, with almost no limits on either vocabulary or sentence structure; second, because it is almost as open-ended with respect to the type fonts and page formats that must be accommodated; and third, because it looks so easy--but has proved so difficult--to find any cheap and simple substitute for spoken language as a means of conveying the printed message to the blind reader. The reading machine problem has been described in greater detail elsewhere (Studdert-Kennedy and Cooper, 1966; Cooper et al., 1969); moreover, its solution is beginning to be realized in a practical way. Field trials of a blind reading service are now under way with the aid of the Prosthetic and Sensory Aids Service of the Veterans Administration and the cooperation of a group of blind veterans at the West Haven Veterans Hospital.

Another potential use for talking computers is in shaping computer-assisted instruction to the needs of the lower grades. The youngsters could learn more easily from a talking informant than from conventional teletype or CRT displays, simply because reading is a skill they have yet to learn. A different use that the mature scholar will one day make of speech from computers is easy reference by telephone to a library's holdings, or at least the portion that is stored in machine-readable form. These, and other, applications of computer-generated speech are now beginning to get the attention they have long deserved. (See Flanagan et al., 1970, for a recent review.)

How do today's uses of voice response differ from the familiar weather and time announcements that telephone companies provide? And why employ computers when magnetic tape and sound track are so much cheaper? An obvious reason is that the simpler devices work well only with restricted message sets and do not extrapolate to more demanding problems. But in the background there is a more subtle reason, namely, that spoken language is

---

*This paper appeared in the 1971 IEEE International Convention Digest (Institute of Electrical and Electronics Engineers, Inc., 1971) 104-105.

+Also, University of Connecticut, Storrs.

by no means simple. This is why even a modest voice-response system needs a great deal of storage, or sophisticated synthesis, or some of each. The nature of the requirement can be seen most clearly in a comparison of the difficulties of outputting a stored message by writing it or by speaking it. Thus, printing a message that is stored in alphanumeric form is a completely straightforward operation and requires only a Teletype machine or lineprinter; by contrast, outputting this same message as speech requires only a little hardware but a lot of know-how and sophisticated programming. To be sure, one can trade memory space for sophistication, but only at a price that is prohibitive for many applications. (It is interesting to note that the inverse problem that computers have in getting an intelligible response from a human is even more difficult. Again, the explanation lies in the inherent complexity of the speech signal.)

THE TALKING COMPUTER AS A LABORATORY HELPER

Let us turn from these general considerations to some specific examples of how one laboratory uses, and plans to use, the voice-response capability that its computer has acquired while engaged in research on reading machines for the blind. We will mention some of the ways in which computer-generated speech is used, then indicate how that speech was produced and what trade-offs were involved in the choice of method.

The preparation of stimulus tapes for experiments on speech perception has been for us a most important use. The computer serves as a versatile informant and can respond appropriately to many variants of the question, "How would you say that?" For experiments on dichotic competition, two messages are needed that are accurately timed with respect to each other. The computer's ability to speak out of both sides of its mouth at once is a great asset. Sometimes the stimulus pairs consist of carefully tailored utterances from a human speaker; sometimes fully synthetic speech is used to get even closer control of intensities, durations, etc. We have made other uses also of the computer's ability to speak in ways that human beings cannot, e.g., more rapidly or slowly, in absolute monotone, adding or omitting formants, and other variants that were needed to test particular hypotheses. Stretched speech has been particularly useful in providing synchronized sound tracks for slow-motion movies of the articulators in action.

Monitoring type-in of information to the computer is another role for voice answer-back. One phase of the work on reading machines required that extensive texts be fed into the computer (for later conversion to speech recordings) in the form of phonetic transcriptions. A keyboard and storage oscilloscope used on-line provide the equivalent of a phonetic typewriter. Typing errors that might otherwise go unnoticed become immediately apparent when automatic voice feedback is provided for each utterance segment or sentence as it is typed.

A cheap remote terminal for inputting programs or data is an obvious variant. If one is copying handwritten programs, there is no real need for a visual display; only a keyboard and loud speaker are needed for type-in and verification on a line-by-line basis.

Confirmation (or correction) of responses in psychological experiments can best be given by voice when the task is a visual one. Thus, in combined dichoptic and dichotic perceptual tests (which need computer control of the stimuli in any case), we expect to use voice responses from the computer to alert and inform the subject.

Scoring test sheets on which subjects have written what they heard in a perceptual test and then entering the data into the computer for processing is an irksome but necessary part of our research. We now have a simple way of hand-scoring the test forms. In one pass, the entries on the form are identified and entered directly into the computer. The scorer has an immediate visual check on his part of the job, and voice answer-back from the computer serves to let him know that the computer correctly understood his written entries.

METHODS FOR GENERATING VOICE RESPONSES; TRADE-OFFS

There are only a few basic methods for getting speech from a computer, but a wide variety of ways to combine them for the best match to a particular task. Usually, the key factor is the form in which the information is stored. One method stores messages as speech wave-forms. This is a simple solution, but it makes heavy demands on memory capacity--roughly 50,000 bits per second of stored speech. The method is most useful when a very limited set of words will suffice for all messages. Thus, for example, the confirmation of subjects' responses or of hand-printed entries needs only a few different responses; we chose waveform storage for these tasks because it is so simple to enter the responses into storage and to use them as output.

Another method suitable for whole messages, or for messages composed of smaller fragments, is to store the control parameters that will enable a synthesizer to generate the speech waveforms. The advantages of this method are considerable: storage requirements are reduced by a factor in the range of 10-50, and there is more flexibility in composing realistic sentences since stress and intonation can be imposed on a composite spectral "skeleton." Disadvantages are that special hardware is needed to synthesize the speech, that the parameters are not related in a simple way to either the speech waveform or the message in its normal written form, and that, in consequence, a substantial amount of processing is required to derive the parameters or to revise them when the message set is to be changed. This is a method of storage we have used in compiling stimulus tapes for psychological experiments.

A third method is synthesis by rule. It is a two-stage process: the first converts a written English sentence into a phonetic transcription, and the second converts the transcription into control parameters for a synthesizer. In many applications, the first conversion will be by-passed since messages in phonetic form require even less storage than in alphabetic form and there are further substantial savings in processing time and memory space. The synthetic speech that can now be generated is still far from perfect, even after more than twenty years of research; nevertheless, it is good enough to be useful to the blind and is the spoken output used in our reading machine research. The advantages of synthesis by rule go well beyond those of synthesis from stored parameters: there

COMPARISON OF METHODS

FOR GENERATING VOICE RESPONSES

Speech Generation	Message Storage	Bit Rate from Message Store
1) Compilation	Waveforms, i.e., voice recordings	40,000 to 60,000
2) Synthesis	Control parameters	1,000 to 4,000
3a) Synthesis by rule	Phonetic symbols	60-100
3b) Synthesis by rule	Alphabetic Text	100-150

COMPARISON OF METHODS
FOR GENERATING VOICE RESPONSES

METHOD	ADVANTAGES	DISADVANTAGES
1) Compilation of voice waveforms (40-60,000 bps)	Simple hardware and software Easy to update messages	Needs very large random-access memory Slow, stilted speech Inflexible sentence structure
2) Synthesis from stored parameters (1-4,000 bps)	Moderate memory requirements Moderate flexibility Fair-to-good speech	Needs hardware synthesizer Parameters not easily derived Messages not easily updated
3a) Synthesis-by-Rule: phonetic symbols (60-100 bps)	Modest memory requirements Flexible Easy to update messages Better speech to be expected as rules are improved	Needs hardware synthesizer CPU time to compute parameters Sophisticated programs
3b) Synthesis-by-Rule: Alphabetic Text (100-150 bps)	SAME AS ABOVE, AND: Even easier to update	SAME AS ABOVE, BUT: More computation More memory (for word lists)

is a further reduction in storage requirements by a factor of 10-20; it is possible to deal with words that are not in the stored vocabulary; and the message set is easy to make or revise, since the entries stored in memory are just the messages themselves, in written form. The disadvantages of synthesis by rule are comparable with those of synthesis from stored parameters: a hardware synthesizer is needed and the programming--after one knows how to do it--is no more complex. However, the demands for processing time are greater, and one cannot hand-tailor the parameters (as one can when they comprise the stored message) to improve the naturalness of the responses. Clearly, though, the future lies with synthesis by rule, as the rules are improved and as more demanding applications are attempted.

SUMMARY

We have tried to explain that the gift of speech--now rare with computers--will one day become very common and very useful. We are only beginning to master the synthesis-by-rule techniques that will make this promise come true, but talking computers can already do a variety of jobs, including the sophisticated one of reading to the blind and the simpler but useful tasks around the laboratory that have been described and illustrated.

REFERENCES

Cooper, F.S., 1963. Speech from stored data. IEEE Convention Record 7, 137-149.

Cooper, F.S., J.H. Gaitenby, I.G. Mattingly, N. Umeda, 1969. Reading aids for the blind. IEEE Trans. Audio and Electroacoustics AU-16, 266-270.

Flanagan, J.L., et al. 1970. Synthetic voices for computers. IEEE Spectrum 7, No. 10, 22-45.

Studdert-Kennedy, M., and F.S. Cooper. 1966. High-performance reading machines for the blind. Proc. Internatl. Conf. on Sensory Devices for the Blind. (London: St. Dunstan's) 317-342.

[The oral version of this paper presented at the 1971 IEEE International Convention (March 22-25, 1971) included a film demonstration of 1) the hand-scoring of answer sheets from psychological tests, 2) the type-in of a phonetic text, with computer monitoring, and 3) on-line modifications being made to synthetic speech. In addition, tape recordings were used to demonstrate page-length passages of text synthesized by rule. In some passages the text was supplied to the computer as a phonetic transcription; in others, the entire process of converting printed text into spoken form had been simulated.]

Audible Outputs of Reading Machines for the Blind*

Franklin S. Cooper, Jane H. Gaitenby, and Ignatius G. Mattingly+
Haskins Laboratories, New Haven

The generation of spoken English from printed text to serve as the output of a reading machine for the blind is the objective of the research at Haskins Laboratories being carried out under a Veterans Administration contract. A series of listening tests using Compiled Speech texts have been made in recent months, and tests have begun of texts recorded in Synthetic Speech. The purpose of these tests was to find out what blind listeners like (and what they will tolerate) in these forms of machine speech.

Compiled Speech and Synthetic Speech are dissimilar approaches to the reading machine output problem. Compiled Speech, an interim approach, was developed because it was an obvious and accessible exploratory strategy. Sentences in Compiled Speech are built up word by word from a prerecorded spoken vocabulary of 7,200 items. The word assembly is done by computer from an input of punched tape that corresponds to a printed text. (See Fig. 1.) Synthetic Speech, on the other hand, is more promising as a long-term approach. It consists of audible sentences which have been generated electronically from (at present) a typed phonetic input. This temporary input method simulates word retrieval by computer from a large dictionary, stored in the computer memory, of the spelled and phonetic forms of words.

Listeners and Recordings

The listeners were veterans attending the Eastern Blind Rehabilitation Center, Veterans Administration Hospital, West Haven, Connecticut. All were male volunteers, acquired as subjects for the tests through the warm cooperation of Mr. George Gillispie, Chief of the Center, and his assistants, of whom Mr. William Kingsley deserves special thanks. There were a total of eleven subjects, most of them in their early twenties. There were eight hour-long tests (each presented to a minimum of two listeners and to a maximum of four).

Thirty-six sample tapes (twenty-seven different texts or versions of texts) were presented in the course of the tests. Listening time per sample ranged from less than a minute to about ten minutes. Four tapes of Synthetic Speech (three different texts) were used in the beginning phase of this part of the study. An attempt was made to control various characteristics of the readings:

1. Rate of the Speech

   The range tested extended from 70 to about 225 wpm. Ten rates within this range were preselected for testing.

---

*Summary prepared for the Bulletin of Prosthetics Research BPR 10-15, Spring 1971.
+Also, University of Connecticut, Storrs.

INTERIM WORD READING MACHINE FOR THE BLIND

PRINT TO AUDIBLE VERBAL OUTPUT USING COMPILED SPEECH

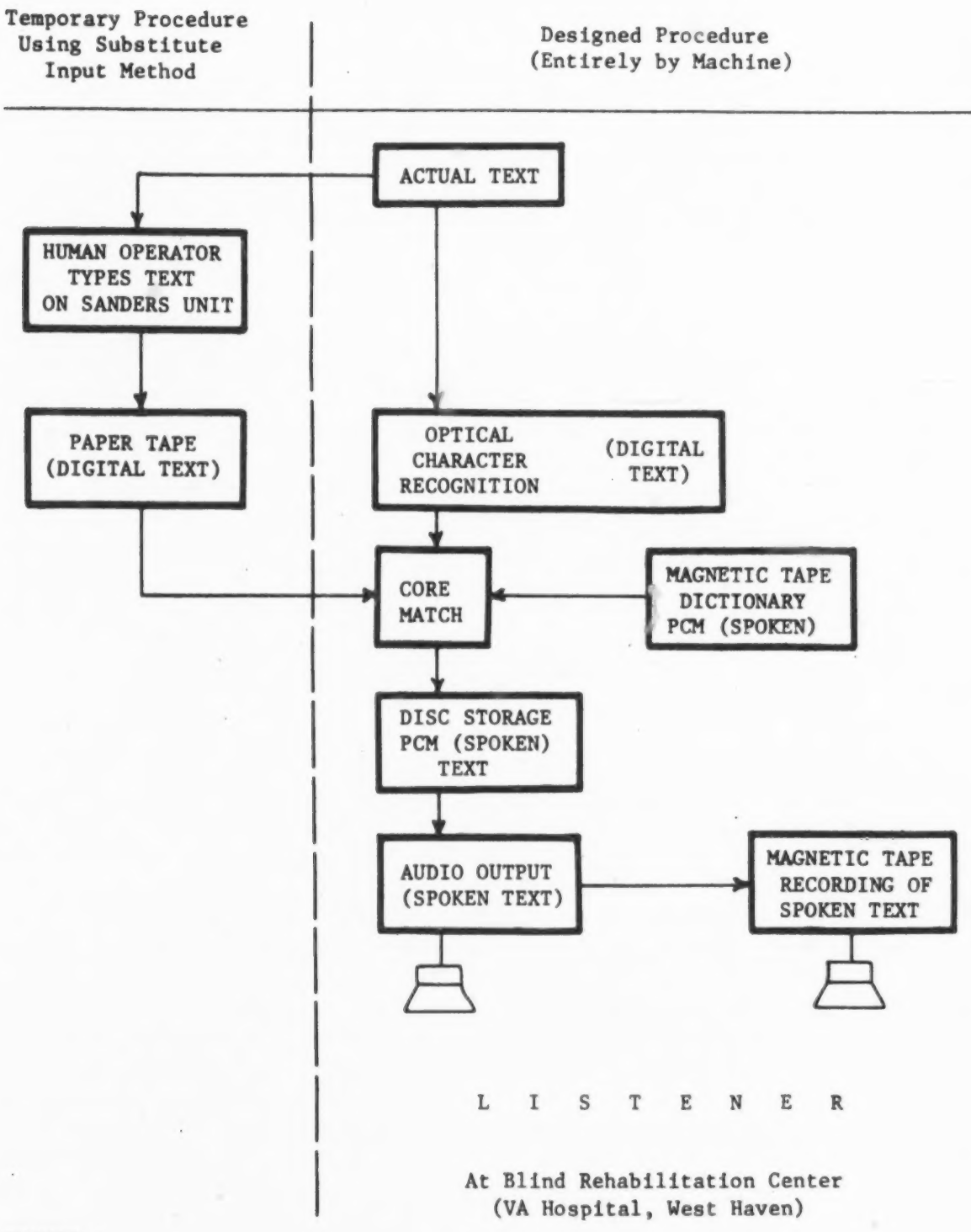

Temporary Procedure Using Substitute Input Method
Designed Procedure (Entirely by Machine)
ACTUAL TEXT
HUMAN OPERATOR TYPES TEXT ON SANDERS UNIT
PAPER TAPE (DIGITAL TEXT)
OPTICAL CHARACTER RECOGNITION
(DIGITAL TEXT)
CORE MATCH
MAGNETIC TAPE DICTIONARY PCM (SPOKEN)
DISC STORAGE PCM (SPOKEN) TEXT
AUDIO OUTPUT (SPOKEN TEXT)
MAGNETIC TAPE RECORDING OF SPOKEN TEXT
L I S T E N E R
At Blind Rehabilitation Center
(VA Hospital, West Haven)

FIGURE 1

2. Type of Rate Manipulation (in Compiled Speech Only)

   a. Time-compressed tape

      Compression is accomplished by dropping regular segments of the speech at brief, regular intervals. Fundamental frequency remains as it was in the original. The speeded recordings were prepared by the Center for Rate Controlled Recordings, University of Louisville, Louisville, Kentucky. The original rate in Compiled Speech was from 100 to about 115 wpm. The amounts of time compression tested were 60, 65, 70, and 75 percent. Discard segments of 10, 15, and 20 msec were tested at each compression.

   b. Capstan-speeded tape

      Here, the output rate is changed by using a different capstan. The fundamental frequency rises proportionately to the increase in rate.

   c. No rate manipulation

      (It should be noted that overall average rates of less than 100 wpm were a function of the amount of spelling that occurred in given texts.)

3. Text (Author and Topic)

   Among these were Dickens, Oliver Twist; Steinbeck, Travels with Charley; Pierce, Waves and Messages; a New York Times sports feature entitled "Hit and Write."

4. Amount of Spelling

   This variable applies only to Compiled Speech, in which all words of the text that are not contained in the prerecorded vocabulary are spelled, letter by letter.

5. Form of Machine Speech Production

   Compiled or Synthetic

The Test Situation

The test sessions took place at the Rehabilitation Center, in whatever room was available. The atmosphere was relaxed and somewhat informal. The investigator began each session with a brief introduction to reading machine research and stressed to the listeners that there were no right or wrong answers to the tests and that, in fact, no answers as such were needed--only candid comments on anything about the tapes that they cared to mention. It was made clear that the purpose of the tests was the eventual improvement of a reading machine output. All the subjects took the task very seriously.

It was thought by the investigator that the listeners' gross reactions--

their free and unprompted comments after each sample--would provide both the broadest and the simplest type of qualitative judgment. The test results given below are a summary of the listeners' spontaneous opinions, as recorded at the time of the tests.[1]

Results with Compiled Speech

At its original recorded rate (100 to 115+ wpm), Compiled Speech has errors of intonation and word-phrase stress and timing. Listener opinions indicate that these deficiencies are more evident in time-compressed tapes when the word rate is either above or below the middle of the range that was tested, i.e., preferred rates were on the range of 150-175 wpm--about normal speaking rates. In capstan-speeded speech, rhythmic and inflectional oddities are most intrusive above 120-135 wpm. With no rate manipulation, Compiled Speech is considered too slow for comfortable listening.

Any spelling in a text, at any of the tested rates, is rejected by the listeners. (One subject said that he was unable to follow spelling even when it was done by a human reader, reading the Morning Telegraph aloud to him.)

In the time-compressed tapes, there was a slight preference for the 20 msec discard interval over the 15 msec discard interval, at all rates. The 10 msec interval was unsatisfactory because "warble effects" were produced in the voice quality, though this may be an artifact of the fundamental frequency of the speaker's voice.

The preferred word rates varied with the subject matter of the text and with the author's style. Also, some topics required more spelling than others. A simple chapter from An Introduction to Sculpture by William Zorach, in which there was very little spelling, was of great interest to the blind listeners, although it was played at the relatively slow rate of 110 wpm. A humorous Saroyan story about a vain young man asking for a screen test (in which there were no spelled words) held their attention through many samples of the same text (played at various rates in time-compressed, capstan-speeded, and Synthetic Speech). Yet two other stories by Saroyan (also without spelling) were deemed so dull, regardless of rate, that the subjects could scarcely bear to listen to them. (One story was largely a dialogue; the other, a monologue.)

The length of the speech sample is another factor in the appraisals. Half a minute is too short an exposure for reasonable evaluation if the topic of the text is unknown and if the tape is begun at a random location in the text. A minimum of one minute seems adequate, no matter what the topic or where the tape is begun, at least if the tape rate is within a reasonable range.

---

[1]It must be pointed out that these were preliminary tests. Although the major aim was to have the machine speech evaluated, further purposes were to zero in on aspects of the tapes that should be controlled in later tests and on aspects of the speech itself that should be improved. (The interactions of the variables is an aspect of note, as the results show.)

The overall evaluation of Compiled Speech that emerged from these tests is that it is acceptable at some rates in either time-compressed or capstan-speeded form, but it is not enjoyable. Spelling is its worst feature. Temporal irregularities are also distressing. Listeners doubted that Compiled Speech (with or without spelling) would be tolerated for extended periods of time.

Preliminary Tests with Synthetic Speech

Four samples of speech synthesized by rule have been tested thus far, and the provisional evaluations represent the views of only six listeners. Nevertheless, responses were generally so positive that it seems useful to include a summary of their reactions.

Synthetic Speech was quite easily understood with exposures as brief as half a minute and at rates ranging from 136 to approximately 225 wpm. It was not necessary to time compress or to capstan speed Synthetic Speech tapes because texts at a variety of rates are easily made within the synthesis-by-rule program itself. (Rates higher or lower than those tested are also readily produced by the program.)

After hearing Synthetic Speech, the listeners' comments dealt mostly with the subject matter of the texts, indicating that the prosody (intonation, etc.) was acceptable, or at least not distracting. This was seldom true of Compiled Speech, which evoked comments dealing chiefly with rate irregularities. The one aspect of Synthetic Speech that was faulted was what the listeners called its "accent"--which some of them actively enjoyed. (One man from inland Maine called it "Bostonian"; another, from Brooklyn, thought it "foreign.")

One listener called Synthetic Speech "Great...really cool!" Such unrestrained approval probably reflects assets that are intrinsic to Synthetic Speech but not to Compiled Speech: no words are spelled; rates are equivalent to, or faster than, normal speech; there are normal transitions between words; intonation is naturalistic and flowing; pauses are reasonable; and rhythm is realistic.

It must be noted, however, that two important liberties were taken in the production of these Synthetic Speech tapes. First, the assumption was made that its stored reference vocabulary contained all of the text words encountered. (In the final machine, as now planned, new or rare words will be generated by rules for syllabification and syllable synthesis.) Second, only one of the three texts depended on rules for parsing, i.e., for stress and juncture assignments; the other two tapes had stress signals supplied by a human operator.

Conclusion

Comparison of the appraisals that have thus far been made of the two forms of machine output indicate that Compiled Speech is effectively rejected in favor of Synthetic Speech. Natural voice quality (commented on approvingly in Compiled Speech when played at its original rate) is one obvious lack of Synthetic Speech. Much of the work now under way is aimed at improvements in the naturalness of Synthetic Speech and at the production of additional and longer tests for a more adequate evaluation of its usefulness as a reading machine output.

PART II: THESIS

TEMPORAL FACTORS IN PERCEPTION OF DICHOTICALLY PRESENTED STOP CONSONANTS AND VOWELS

Emily Flora Kirstein, A.B.

Goucher College, 1963

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

at

The University of Connecticut

1971

ACKNOWLEDGMENTS

This work grew out of the ongoing program of research on dichotic listening at Haskins Laboratories. I am especially grateful to Dr. Donald Shankweiler and Dr. Michael Studdert-Kennedy, who discovered the "lag effect," to Dr. Franklin Cooper, who generously made available the excellent research facilities at Haskins Laboratories, and to my major adviser, Dr. Alvin Liberman, who persuaded me that the "lag effect" was interesting and worthy of further study. This research would not have been possible without the encouragement of all of these people and many others at Haskins Laboratories and at the University of Connecticut.

Many thanks are due to Dr. Liberman and Dr. Shankweiler for their patient and insightful readings of the manuscript, to Dr. Kenneth Ring for many years of encouragement, and to Agnes McKeon for assistance with the illustrations. I would also particularly like to thank my parents, who supported this work in innumerable ways, and especially my mother, Judith Kirstein, who typed the manuscript.

This research was supported in part by grants to Haskins Laboratories from the National Institute of Child Health and Human Development (HD-01994) and from the National Institutes of Health (General Research Support grant FR-5596). Tabulation of the data was greatly aided by access to the University of Connecticut Computer Center made possible by a grant from the National Science Foundation (GJ-9).

TABLE OF CONTENTS

Acknowledgments i

Table of Contents iii

List of Tables iv

List of Figures v

INTRODUCTION 1

EXPERIMENT 1. A COMPARISON OF THE EFFECTS OF MONOTIC AND DICHOTIC PRESENTATION ON THE PERCEPTION OF TEMPORALLY OVERLAPPED STOP CONSONANT-VOWEL SYLLABLES 7

EXPERIMENT 2. SELECTIVE LISTENING FOR DICHOTICALLY PRESENTED STOP CONSONANTS 27

EXPERIMENT 3. PERCEPTION OF STOP CONSONANTS AND VOWELS IN DICHOTICALLY PRESENTED CV SYLLABLES 44

EXPERIMENT 4. EFFECTS OF DELAY BETWEEN EARS ON THE PERCEPTION OF DICHOTICALLY PRESENTED ISOLATED STEADY-STATE VOWELS . . 53

SUMMARY 62

BIBLIOGRAPHY 66

LIST OF TABLES

I Syllable pairs included on the stimulus tape in Experiment 1 . . . 9

II Syllable pairs included on the stimulus tape in Experiment 3 . . . 46

III Individual differences in the magnitude of the lag effect correlated across the C, CV-C, and CV-V conditions (Spearman rank correlation coefficients) 51

IV Individual differences in the magnitude of right-ear effect correlated across the C, CV-C, and CV-V conditions (Spearman rank correlation coefficients) 51

V Comparison of the C, CV-C, CV-V, and isolated vowel conditions for naive subjects. The incidence of lag effects, the incidence of right-ear effects, and the relation between the ear effect and lag effect 58

LIST OF FIGURES

1 Spectrograms of the nine stimulus syllables 8

2 Mean percent correct responses as a function of the interval between syllable onsets for the monotic and dichotic conditions . . 12

3 Mean percent correct responses as a function of stimulus lag or lead time for the monotic and dichotic conditions 13

4 Mean percent correct first responses as a function of stimulus lag or lead time for the monotic and dichotic conditions 14

5 Mean percent correct second responses for the dichotic and monotic tests 16

6 Frequency distributions of lag effect scores, (Leading-Lagging)/(Leading+Lagging), for the dichotic and monotic conditions 18

7 Frequency distribution of ear effect scores, (R - L)/(R + L), based on first responses in the dichotic condition 19

8 Comparison of left-ear lag and right-ear lag trials in the dichotic condition 20

9 Percent correct first responses by ear on the dichotic test for subject SV 22

10 Percent correct first responses by ear on the dichotic test for subject BR 23

11 Mean percent correct first responses by ear comparing the same delay conditions for the two ears 24

12 Comparison of the form of the lag effect function for the two ears after a displacement of the left-ear curve 20 msec along the x-axis 25

13 Accuracy of selecting responses by ear and by temporal order as a function of the time between syllable onsets 30

14 Accuracy of selecting responses from the left and right ears . . . 31

15 Accuracy of selecting lagging and leading stimuli 32

16 Mean percent correct responses on the ear-monitoring and temporal order tasks as a function of ear and relative onset time of the correct syllable 33

17 Mean percent intrusions on the ear-monitoring and temporal tasks as a function of the ear and relative onset time of the intruding syllable 34

18 Changes in the magnitude of the right-ear effect and lag effect as a function of interaural delay time for the ear-monitoring and temporal order tasks 36

19 Comparison of the frequency distributions of lag effect scores obtained with ear monitoring and clarity judgments 37

20 Comparison of the frequency distributions of ear effect scores obtained with ear monitoring and clarity judgments 41

21 Hypothetical lag effect function with an assumed peak at 60 msec delay and the right- and left-ear curves which would result if a linearly decreasing right-ear effect were added orthogonally to the lag effect at each delay interval 43

22 Mean percent correct first and second responses as a function of the interval between stimulus onsets for conditions C, CV-C, and CV-V 48

23 Mean percent correct responses for lagging and leading stimuli for conditions C, CV-C, and CV-V 49

24 Mean percent correct first responses corresponding to lagging and leading stimuli for conditions C, CV-C, and CV-V 50

25 Mean percent correct first responses by ear for "naive" and "experienced" subjects on the isolated vowel test 55

26 Scatter plot showing the relation between an individual's lag effect score (y-axis) and ear effect score (x-axis) for the isolated vowel test 56

27 Mean percent correct first responses as a function of lag or lead time for seven subjects on the C, CV-C, CV-V, and isolated vowel tests 59

INTRODUCTION

The subject of this thesis is a newly discovered perceptual phenomenon which has been termed the "lag effect." The lag effect was first observed in a dichotic listening[1] experiment where stop consonant-vowel syllables[2] contrasting in the stop consonant were delivered to opposite ears with slight differences in the time of arrival of the syllables at the right and left ears. When listeners were asked to identify the competing stop consonants, they were more often correct in identifying the delayed syllable than the one which was first to arrive (Studdert-Kennedy, Shankweiler, and Schulman, 1970; Lowe, Cullen, Thompson, Berlin, Kirkpatrick, and Ryan, 1970; Porter, Shankweiler, and Liberman, 1969). The term "lag effect" refers to this advantage in recall accruing to the syllable at the delayed ear. This advantage is observed regardless of whether the delayed ear is the right ear or the left ear.

We still do not know what experimental conditions are required in order to elicit the lag effect. It has been established, however, that two sorts of competition are needed. There must be competition between stimuli and also competition between ears. When the temporally overlapped CV syllables were made to compete at the same ear, the result was the opposite of that observed with dichotic presentation. With monotic presentation[3] leading stop consonants were more accurately identified than lagging stop consonants (Studdert-Kennedy et al., 1970; Lowe et al., 1970).

Of the two effects--the dichotic lag effect and the monotic lead effect--the dichotic effect excited greater interest because it clearly had its origin in competition arising within the central nervous system. The monotic lead effect, on the other hand, could plausibly be attributed to masking at the peripheral receptor.

In order for us to make a more precise statement about the locus of the lag effect, more information is required. The general goal of the research described in this thesis was to provide data which would aid in pinpointing more exactly the stage in the processing of the stimuli at which the lag effect originates.

It is possible that the lag effect is a phenomenon of auditory perception. If this is so, then it should be possible to obtain the lag effect with a fairly wide assortment of auditory stimuli--nonspeech stimuli as well as speech stimuli. However, Studdert-Kennedy et al. (1970) and Porter et al.

---

[1] Dichotic presentation is an experimental method in which different auditory stimuli are delivered simultaneously to opposite ears.

[2] The stimuli were the syllables [pa], [ta], [ka], [ba], [da], [ga].

[3] Monotic presentation is the presentation of different stimuli simultaneously at the same ear.

(1969) considered the lag effect to be a more circumscribed phenomenon. They proposed that the lag effect occurs only for speech sounds and that the effect reflects some property of speech processing mechanisms. If this idea is correct, then an understanding of why the lag effect occurs could greatly increase our knowledge about the nature of the processes involved in speech perception.

The notion that the lag effect is a speech perception phenomenon seems reasonable from a theoretical point of view. There is considerable evidence that there exist distinct perceptual systems for processing speech and nonspeech stimuli (Liberman, Cooper, Shankweiler, and Studdert-Kennedy, 1967). Research on speech perception has demonstrated that the relationship between the acoustic speech signal and the perceived phonological structure of the message is exceedingly complex, much more complex than the relation between stimulus and percept in the case of nonspeech auditory stimuli. One complexity peculiar to speech perception has to do with the problem of segmentation of the acoustic signal. The phonetic message is perceived as a string of segments; for example, the word "mat" is perceived as having three segments or phones, [m], [æ], and [t]. However, these segments are not physically marked in the acoustic signal. Rather, the signal is essentially continuous, and the minimal units into which the signal might be segmented on some purely physical basis are more on the order of syllables than individual phones. In addition to the lack of segmentation, another complexity in the relation between the acoustic speech signal and percept is the variability in the acoustic representation of a particular speech sound depending on the surrounding phones. For example, in the word "mat," the features of the acoustic signal which distinguish [m] from [n] or [t] from [k] vary greatly depending on whether the vowel is [æ] or some other vowel, and in rapid speech the acoustic cues for the vowel are themselves altered by the identity of the initial and final consonants. Thus, the entire syllable is recoded into a single acoustic unit, and the task of the listener is one of decoding the signal in order to extract the string of phonetic segments (Liberman, 1971). This parallel transmission of sequential information appears to account for the unique efficiency of speech as a vehicle for transmitting language (Liberman, Cooper, Studdert-Kennedy, Shankweiler, and Harris, 1966), but special apparatus would seem necessary for the recognition of speech sounds.

One of the most convincing sources of evidence of the existence of distinct speech and nonspeech modes of perception is research indicating that speech and nonspeech sounds are analyzed in opposite brain hemispheres. It has long been known that in man the left cerebral hemisphere is more important than the right cerebral hemisphere for the perception and production of language. This conclusion was based originally on studies of the effects of unilateral brain injury and on diagnostic procedures like the paralysis of one hemisphere by the injection of intracarotid sodium amytal (Wada and Rasmussen, 1960) and the electrical stimulation of the cortex (Penfield and Roberts, 1959). More recently an experimental technique has become available for investigating lateralization of function in normal people. Use of this technique has led to the conclusion that the analysis of speech sounds takes place predominantly in the left hemisphere along with other linguistic functions, while analysis of nonspeech auditory stimuli takes place predominantly in the right hemisphere.

These conclusions are based on the results of certain dichotic listening experiments. Kimura (1961a,b) discovered that when digits are presented in

competition at opposite ears, the digits presented at the right ear are more accurately identified than those simultaneously presented at the left ear. A right-ear superiority has also been demonstrated for dichotically presented meaningful words other than digits (Dirks, 1964; Curry and Rutherford, 1967; Borkowski, Spreen, and Stutz, 1965) and for nonsense words (Kimura, 1967; Curry and Rutherford, 1967). It is now generally accepted that the right-ear advantage in report of dichotically presented verbal material is a consequence of the left-hemispheric specialization for language. Kimura (1961b, 1967) attributed the ear asymmetry to the fact that the contralateral pathway from ear to cortical processing areas is stronger than the ipsilateral path. The superiority of the crossed path would give stimuli at the right ear a perceptual advantage over stimuli at the left ear when the stimuli are competing for the linguistic processing areas in the left hemisphere.

This interpretation of the ear effect, which links the right-ear advantage to left-hemispheric language "dominance," received strong support from the finding that people with reversed (i.e., right-hemisphere) language representation as assessed independently by the sodium amytal test also had a reversed (i.e., left-ear) advantage on the dichotic digits task (Kimura, 1961a). Further support came from the finding that most people show a left-ear advantage when the stimuli are dichotically presented nonspeech sounds like melodies, sonar signals, or environmental sounds (Kimura, 1964; Curry, 1967; Chaney and Webster, 1966; Spreen, Spellacy, and Reid, 1970; Darwin, 1969). The left-ear effect for dichotically presented nonspeech sounds implies that the right hemisphere is of greater importance than the left for nonspeech auditory perception. The results of research on effects of right and left temporal-lobe damage also support the view that the right hemisphere is specialized for processing nonspeech material (Milner, 1962; Shankweiler, 1966).

The fact that the direction of ear asymmetry is different for speech and nonspeech auditory stimuli does not necessarily mean that the identification of speech sounds takes place in the left hemisphere. Left-hemispheric specialization may pertain only to the higher (semantic and syntactic) levels of language. The actual analysis of the acoustic signal and the identification of the sounds might take place predominantly in the right hemisphere along with other auditory analysis. It was soon discovered that neither grammatical structure nor meaningfulness was a necessary condition for obtaining the right-ear effect. The right-ear advantage could be reliably obtained when the dichotically presented stimuli were CV or CVC syllables contrasting between ears in the initial or final consonant. Report from the right ear is more accurate than report from the left ear for competing stop consonants (Shankweiler and Studdert-Kennedy, 1966, 1967; Studdert-Kennedy and Shankweiler, 1970), fricatives (Darwin, 1971b), or liquids (Haggard, 1969). The finding of a right-ear superiority when single phonetic segments differ between ears and a left-ear superiority when nonspeech stimuli differ between ears is strong evidence in support of the distinction between speech and nonspeech perceptual modes. This result also ties the speech perception mode closely to the remainder of the linguistic system.

If there are distinct speech and nonspeech modes of perception, then it should be possible to discover perceptual phenomena which occur for speech stimuli but not for nonspeech stimuli. The dichotic right-ear effect is one such phenomenon. Another phenomenon which apparently is found only with speech

sounds is the phenomenon of categorical perception. Categorical perception is demonstrated in an experiment where listeners are required to discriminate small variations along some acoustic dimension which serves as an acoustic cue for phonetic classification of the stimuli. If the stimuli are encoded speech sounds like stop consonants, listeners are unable to make auditory judgments about stimuli which are acoustically different but fall within the same phonetic class. That is, listeners perceive the phonetic category and fail to discriminate the within category acoustic variations. If the same acoustic dimension is removed from the speech context, these auditory distinctions are readily made (Eimas, 1963; Liberman, Harris, Kinney, and Lane, 1961; Mattingly, Liberman, Syrdal, and Halwes, 1971).

Both categorical perception and a dichotic right-ear advantage are reliably obtained when the stimuli are the stop consonants. These same stimuli were the ones used in the experiment in which the lag effect was discovered. In fact, when syllables contrasting in the stop consonant are presented dichotically with slight differences in onset time, both a right-ear advantage and a lag effect are observed. They appear to be separate phenomena superimposed on each other. Since it is known that the right-ear effect and categorical perception are specifically speech perception phenomena, it would seem reasonable to suppose that the lag effect might be a speech perception phenomenon as well.

In addition to the theoretical considerations which might lead one to suppose that the lag effect is a speech perception phenomenon, there is also some empirical evidence relevant to this hypothesis. In order to test directly the notion that the lag effect is a speech perception phenomenon, one would like to compare perception of dichotically presented time-staggered speech and nonspeech stimuli. This experiment has not yet been performed. However, Massaro (1970) performed a similar experiment with tones and reports a phenomenon similar to the lag effect. Massaro (1970) delivered a 20-msec test tone to one ear and a longer masking tone to the other ear. When the onset of the masking tone followed the onset of the test tone by 20 to 350 msec, the masking tone interfered with recognition of the test tone. However, no interference was observed when the masking tone preceded the test tone.

The phenomenon described by Massaro is similar to the lag effect, but it differs from the lag effect in one respect which seems significant. This difference is in the time course of the two phenomena. For the dichotic tones, the interference was greatest with the shortest delays between stimulus onsets and decreased monotonically with longer delays. For the dichotic stop consonants, the lag effect in the data of Studdert-Kennedy et al. (1970) was visible with a 10-msec difference in stimulus onsets, but the interference in report of the leading stop increased with longer delays. The lag effect was greatest with a delay of about 50 msec between syllables and then decreased with still longer delays. The effect was still visible with a delay of 120 msec, but with delays of 180 msec or more, the lag effect is apparently no longer in evidence (Berlin, Loovis, Lowe, Cullen, and Thompson, 1970). In the Massaro (1970) experiment the interference persisted to a certain extent even with delays as long as 350 msec between stimulus onset. Because of the difference in the time course of the interference effects observed with speech and nonspeech, it seems unlikely that the identical processes underlie the two phenomena.

Also, it is evident from other research that the phenomenon described by Massaro (1970) is not in itself a completely general phenomenon of auditory perception of dichotic stimuli. Darwin (1971a) opposed stop consonant-vowel syllables at one ear to bursts of noise of the same duration at the opposite ear. He found that when syllables were opposed to noise, the effect of delay between ears was the opposite of that observed when syllables were competing against other syllables. Syllables were better identified when they preceded the noise onset, although they were better identified when they followed the onset of a competing syllable.

If the lag effect and the "retroactive interference" effect described by Massaro (1970) were both instances of the same phenomenon which occurs when auditory stimuli are presented with temporal delays between ears, then one would expect to observe the same effect when speech stimuli are opposed to noise, since both the syllables and the noise are auditory stimuli. Darwin's (1971a) result would seem to suggest that the lag effect arises not merely out of competition between stimuli at the two ears but that, in addition, the stimuli must be competing for the same processor. Stop consonants presented at both ears give a lag effect. Dichotically presented tones give a similar effect. But stop consonants opposed to noise do not give the effect. Thus, even if there is an effect with nonspeech stimuli which is similar to the lag effect, the fact that such interference is not obtained when one of the stimuli is a syllable and the other a nonspeech sound implies that a distinction between speech and nonspeech is involved in these phenomena.

Certain other experiments would seem to implicate speech processing in the lag effect. These experiments indicate that the lag effect is sensitive to the linguistic compostion of the competing syllables. Day (1968, 1969) presented dichotically words differing between ears in the initial consonants. The word at one ear began with a stop consonant while the word at the other ear began with a liquid. The relative onset time of the two words was varied. When asked to report what they heard, about half the subjects reported hearing one or both of the stimulus words on each trial. For these subjects, there was apparently a preference for the lagging stimulus (Day, personal communication). However, the remaining subjects reported hearing a single stimulus which was in fact a fusion of the two stimulus words. For example, when "back" was delivered to one ear and "lack" was delivered to the other ear, these subjects reported hearing "black" regardless of whether the stop led the liquid or the liquid led the stop.

While nearly all subjects have a lag effect with stop-stop competition, at least half the subjects did not show a lag effect with stop-liquid combinations but gave a "fusion" response instead. Stops and liquids are rather similar classes of sounds acoustically, so if the lag effect were thought to arise in the purely auditory analysis of the stimuli, it would be difficult to explain why the effect is so much more frequent with stop-stop than with stop-liquid combinations. The explanation may lie in the linguistic difference between the stop-stop and stop-liquid conditions. Stop-stop sequences never occur in syllable-initial position in English, but initial stop-liquid clusters are common. It is only in the latter condition that fusion responses are possible. It may be that a decision is made whether to combine the dichotic stimuli or to treat them as competing stimuli. Such a decision would have to be made at a rather high level of analysis,

after a partial phonetic analysis has revealed whether the stimuli from the two ears are capable of being fused. This reasoning would place the origin of the lag effect at a stage in processing which follows considerable phonetic analysis of both syllables.

The lag effect may be one of a class of perceptual interference effects which occur when stimuli are in competition for the same central processor. The processors involved may be quite specific--processors of speech sounds, or perhaps even certain classes of speech sounds, and processors of nonspeech sounds. When different processors are required as in the case where speech and noise were presented to opposite ears, there is no evidence of the lag effect. It should also be pointed out that the nonspeech experiment of Massaro (1970) did not use precisely the same experimental technique as that used with the stop consonants, and more research is clearly needed to determine whether there is a lag effect with competing nonspeech stimuli.

The four experiments described in this thesis are all concerned with the dichotic lag effect. The major goal was to establish the level of processing at which the lag effect originates and, specifically, to evaluate the notion that the lag effect is related to phonetic recognition processes. In this connection, the relation between the lag effect and the dichotic right-ear effect is of particular interest. Since it is known that the right-ear effect is obtained only for speech stimuli, it would be expected that some relation between the two effects would be seen if the lag effect also had its origin in speech decoding processes.

Other questions dealt with in the thesis will be mentioned briefly. One problem was to determine whether the lag effect depends on any particular strategy of report. If the lag effect proved to be independent of the method of recall, this would be strong evidence that the effect is a genuine perceptual phenomenon. If, on the other hand, a particular recall procedure were required, one might conclude that the lag effect is related to the organization of responses rather than to perception.

Finally, questions were raised about the subjective experience of people listening to syllables competing at the two ears. Can listeners hear both stimuli or are they aware of only one? Can they tell which stimulus is arriving at each ear? Can they tell which of the syllables is the first to arrive and which is delayed?

EXPERIMENT 1. A COMPARISON OF THE EFFECTS OF MONOTIC AND DICHOTIC PRESENTATION ON THE PERCEPTION OF TEMPORALLY OVERLAPPED STOP CONSONANT-VOWEL SYLLABLES

Experiment 1 was intended to confirm the reports of Studdert-Kennedy et al. (1970) and of Lowe et al. (1970) of differences between the effects of monotic and dichotic presentation on the perception of competing temporally staggered stop consonants. In those experiments the stimuli used were the stop consonant-vowel syllables [ba], [da], [ga], [pa], [ta], and [ka] presented in pairs with delays of 5 to 120 msec between the onsets of syllables within a pair. Both studies showed that when the two syllables arrived at opposite ears, the lagging stop was more accurately reported than the leading stop, but when the syllables arrived at the same ear, the leading stop was more accurately reported than the lagging one.

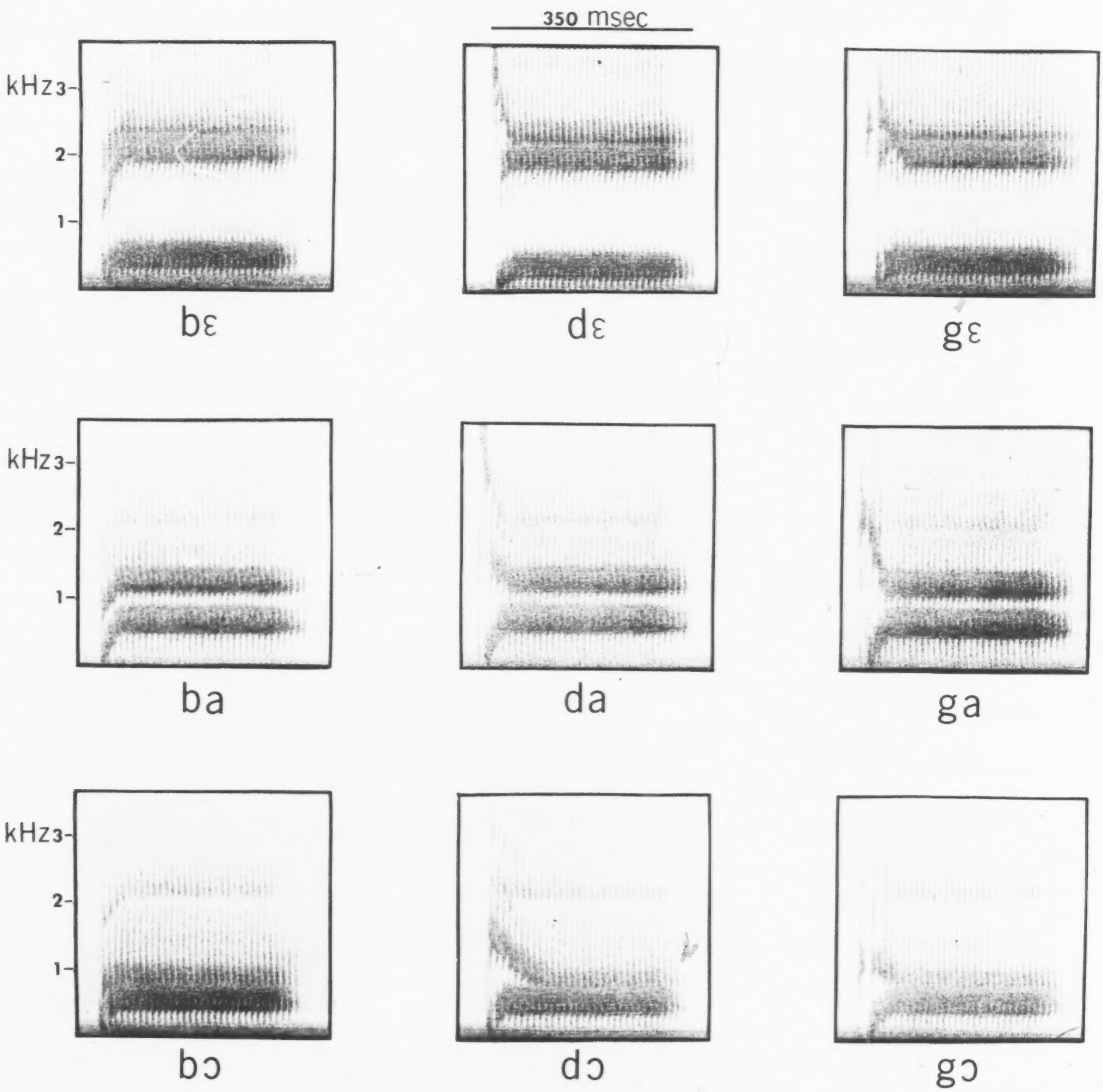
Experiment 1 repeated the comparison between the dichotic and monotic conditions using a somewhat altered stimulus set. The stimulus set was expanded to nine syllables--[bɛ], [dɛ], [gɛ], [ba], [da], [ga], [bɔ], [dɔ], [gɔ]--formed by combining each of the three voiced stop consonants [b], [d], and [g] with each of three vowels, [ɛ], [a], and [ɔ]. Although the set of syllables was enlarged over that used in previous experiments, the response set was reduced to three rather than six possibilities because competing syllables always shared the same vowel. The subjects needed only to identify the stops and could ignore the vowels.

Experiments 2, 3, and 4 made use of the same stimulus set as Experiment 1. Each of these later experiments evaluated the influence of some deviation from the task described in Experiment 1. The results of Experiment 1 thus form a base line against which the later findings can be compared.

Method

Stimuli. Preparation of the stimulus syllables and of the test tape was accomplished with the aid of the computer facilities at Haskins Laboratories (Cooper et al., 1971). Nine syllables--[bɛ], [dɛ], [gɛ], [ba] [da], [ga], [bɔ], [dɔ], [gɔ]--were generated on the Haskins Laboratories computer-controlled parallel resonance speech synthesizer. Spectrograms of these syllables are shown in Figure 1. Each syllable is a 350-msec three-formant pattern. The rapidity of the changes in formant frequency (transitions) in the initial portion of each syllable is the acoustic cue which distinguishes stop consonants as a class from other classes of speech sounds. These transitions occupy only the first 45 to 70 msec of any syllable. The direction and extent of the second and third formant transitions is the primary acoustic cue for the place of articulation of the stops, while the rising first formant transition, which is constant among the nine syllables, is characteristic of voiced stops as opposed to voiceless stops or nasals. The formant transitions are followed by steady-state patterns with formant frequencies and amplitudes appropriate for each vowel. The steady-state formant frequencies for a particular vowel are independent of the place of articulation of the accompanying stop,

Spectrograms of the Nine Stimulus Syllables



Note: The spectrographic representation of an acoustic signal shows intensity by the darkness of the display, frequency (in kHz) by position on the y-axis, and time by position in the x-axis.

Fig. 1

but the direction and extent of the formant transitions for a particular stop vary depending on which vowel follows.

The intelligibility of these syllables was assessed by asking four people with no previous exposure to synthetic speech to identify the syllables in a 180-trial randomized sequence. In the 720 trials only one error was made in identifying the vowels of the syllables. The stop consonants in the syllables [ba], [da], [ga], [bɛ], [gɛ], [dɔ], and [gɔ] were identified correctly 100 percent of the time. On one occasion [dɛ] was heard as [gɛ]. The syllable [bɔ] was often heard as [gɔ]; this error was idiosyncratic in that some subjects consistently made the error while others always identified the sound as it was intended. The overall percent correct identification of 95 percent was considered adequate for the dichotic experiments.

Preparation of the test tape. A two-channel tape was assembled under computer control using the PCM system (Cooper and Mattingly, 1969). The PCM program allows one to specify a random order of pairs of syllables and affords precise control over the interval between the onsets of syllables on the two channels. Using this program, the syllables were first stored on a disc file and then recalled in pairs for recording onto a two-channel tape.

The tape contained eighteen different pairs of syllables--all possible pairings of the nine syllables in which stop consonants differed between channels while vowels were shared (for example, pairs like [ba]-[da] and [gɔ]-[dɔ]). The eighteen pairs of syllables are listed in Table I. For any pair the

TABLE I. Syllable pairs included on the stimulus tape in Experiment 1.

Syllable on Channel 1	Syllable on Channel 2	
BE	DE	Each of these combinations occurs once with a channel 1 lead of 10, 30, 50, 70, and 90 msec and with a channel 2 lead of 10, 30, 50, 70, and 90 msec.
DE	BE	
BA	DA	
DA	BA	
BO	DO	
DO	BO	
BE	GE	
GE	BE	
BA	GA	
GA	BA	
BO	GO	
GO	BO	
DE	GE	
GE	DE	
DA	GA	
GA	DA	
DO	GO	
GO	DO	

The letters E, A, and O represent the phonetic qualities [ɛ], [a], and [ɔ], respectively.

syllable onset at one channel was delayed relative to the other channel by 10, 30, 50, 70, or 90 msec. Each of the eighteen pairs occurred twice at each delay interval, once with channel 1 being the delayed channel, and once with channel 2 delayed. The tape contained 180 stimuli--18 pairs x 5 delays x 2 channels--randomly ordered in a 180-trial tape. Each trial consisted of the presentation of a single pair of syllables. The intertrial interval was 6 seconds for most trials, with 10-second pauses inserted between blocks of ten trials and a 30-second pause after 90 trials.

Testing procedure. The same tape was used for both the dichotic and monotic presentation conditions. For the dichotic test each tape channel was delivered to a different ear. For the monotic test the two channels were mixed electronically and presented to the same ear. In both the dichotic and monotic tests each subject listened to the tape twice within a test session, with the headphones physically reversed on the second run. This gave a total of 360 trials for each subject, 72 trials at each of the five delay intervals. The headphone reversal is an important control on dichotic tests where it is necessary to balance stimulus conditions over the two ears. For the first 180 trials on the dichotic test the subjects always received channel 1 at the left ear and channel 2 at the right ear. For the second 180 trials, the left ear received channel 2 and the right channel 1. For the monotic test the mixture of the two channels was played to the left ear for the first 180 trials and to the right ear for the second 180 trials.

The subjects were tested in groups of one to six people in a quiet room. The tape was played on a General Radio tape deck (Type 1525) into a "listening station" designed and built by D. Zeichner of Haskins Laboratories for group dichotic experiments. The listening station is a two-channel amplifier with multiple outputs for headphones. The output level of the signal from each channel can be adjusted in 1 db calibrated steps. The listening station can be used to present stimuli monaurally to either ear, monotically (both channels mixed electronically and presented to the same ear), and dichotically (a different channel to each ear).

The stimuli were presented at a comfortable listening level (about 75 db at each ear) over headphones (Grason-Stadler TDH-39-300Z). The intensity of the output from the two channels was equated as closely as possible by the use of calibration signals laid down with the same intensity on the two channels at the time of recording of test tape.

Instructions to subjects. As far as possible identical instructions were given for the monotic and dichotic tests. The subjects were told that they would receive two different syllables on each trial, that the syllables would always differ only in the consonant, and that only the consonants "B," "D," and "G" would be used. The task was to identify both consonants on each trial and then to decide which of the two sounded clearer. The clearer of the two sounds was to be recorded in the first column of the answer sheet (first responses) and the less clear was to be recorded in the second column (second responses). Both columns were to be filled in with different responses even if it was necessary to guess. The subjects were told that some trials would be more difficult than others, but they were not told that relative onset time was a variable in the test.

Subjects. All subjects were students in the introductory psychology classes at the University of Connecticut. Their participation in these experiments fulfilled part of their course requirement. All subjects were native speakers of English, were right handed, and had no known hearing defect.

Separate groups of subjects were run on the dichotic and monotic tests. Each group contained twelve subjects.

Results

The listeners had two tasks to perform on each trial--to identify both consonants and to indicate which was the clearer by recording the clearer consonant in the first column (first response) and the less clear consonant in the second column (second response). A response is considered correct if it corresponds to either of the stimulus consonants presented on that trial.

Accuracy of identification. Many more errors in identification were made on second responses than on first responses. On the dichotic test 94.1 percent of all first responses were correct while only 66.6 percent of all second responses were correct. The monotic test did not differ significantly from the dichotic test with respect to accuracy of performance. For the monotic test 90.4 percent of all first responses and 65.6 percent of all second responses were correct.

Figure 2 shows the accuracy of identification of stops in syllable pairs with temporal offsets of 10, 30, 50, 70, or 90 msec. First responses were nearly always correct and the accuracy of first responses was not systematically affected by the amount of time separating the onsets of the competing syllables. The ability of subjects to report both stops correctly was, however, influenced by the delay between syllables. Second responses became more accurate with longer delays for both the dichotic and monotic tests.

Effects of lag or lead on the accuracy of identification. Figure 3 shows the effect of lag or lead time on the accuracy of identification. On the monotic test leading stops were more often correctly reported than lagging stops, but with dichotic presentation lagging stops had an advantage over leading stops. The percentage of correct responses expected by chance in this task is 67 percent. When monotically presented syllables were separated by 10 msec, the lagging stop was reported no more frequently than would be expected by chance, and identification of lagging stops did not rise above chance level until the syllables were separated by 70 msec. Leading stops, in contrast to lagging stops, were accurately reported at all delays on the monotic test. On the dichotic test the suppression of the leading syllables was not so great as the suppression of lagging syllables on the monotic test. Identification of leading stops never fell to chance level, but there was an advantage for lagging syllables at all delay intervals.

Analysis of first responses. The subjects were instructed to indicate which of the stops sounded clearer by recording the clearer stop in the first column of the answer sheet. Figure 4 shows the percentage of trials on the two tests in which lagging and leading stimuli were given as the first response.

Mean Percent Correct Responses as a Function of the Interval Between Syllable Onsets for the Monotic and Dichotic Conditions

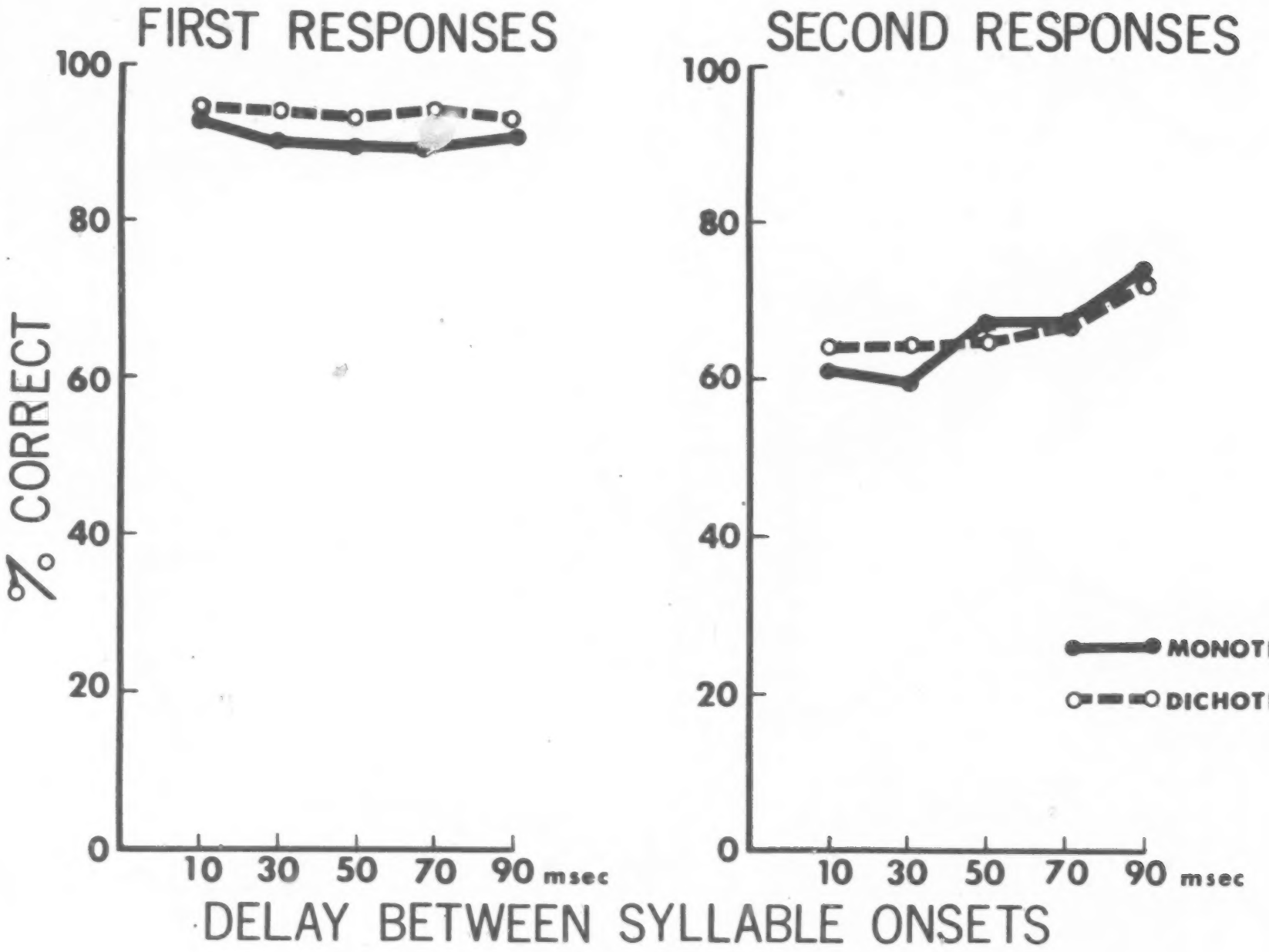


Fig. 2

Mean Percent Correct Responses as a Function of Stimulus Lag or Lead Time for the Monotic and Dichotic Conditions

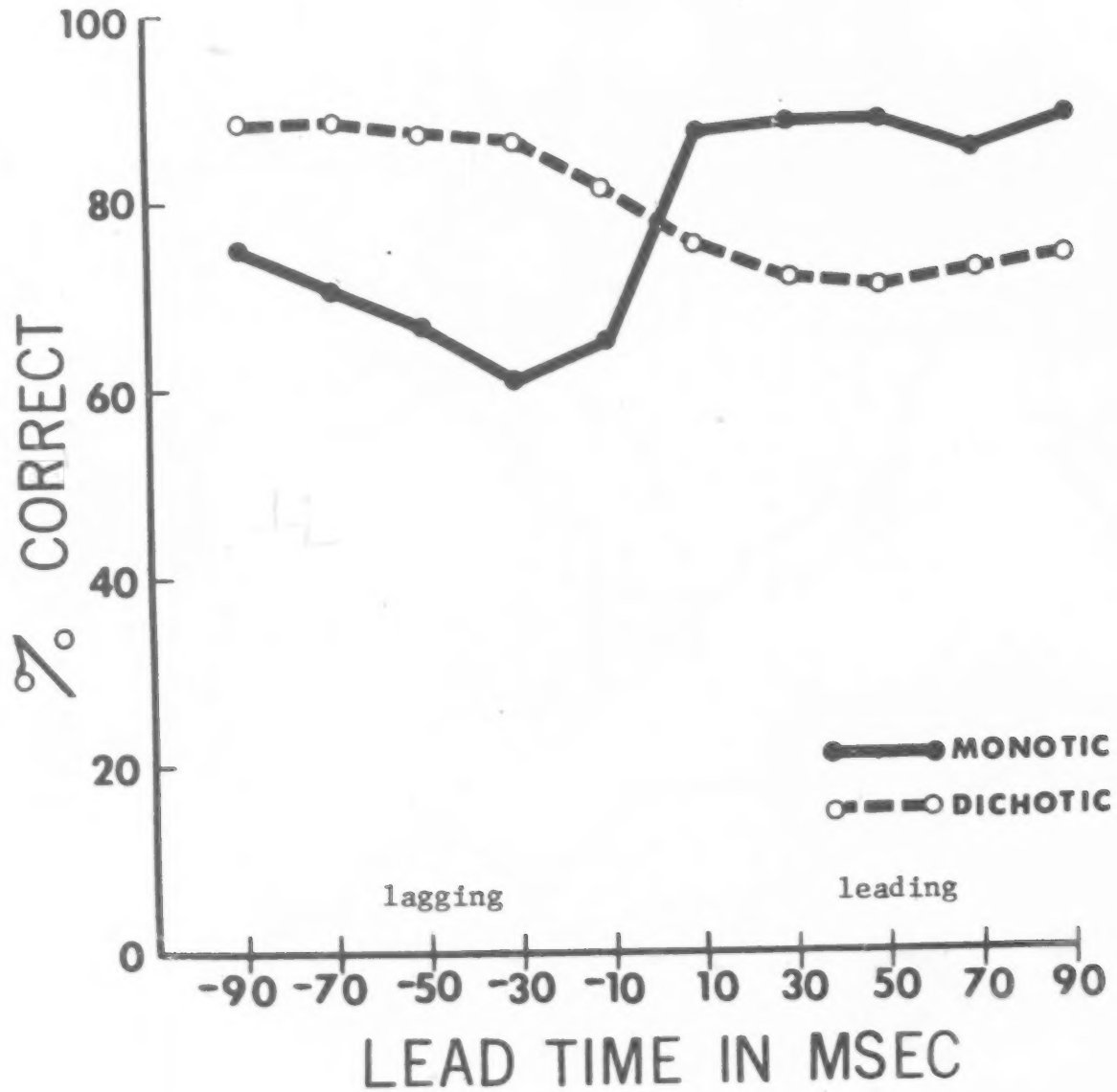


Fig. 3

Mean Percent Correct First Responses as a Function of Stimulus Lag or Lead Time for the Monotic and Dichotic Conditions

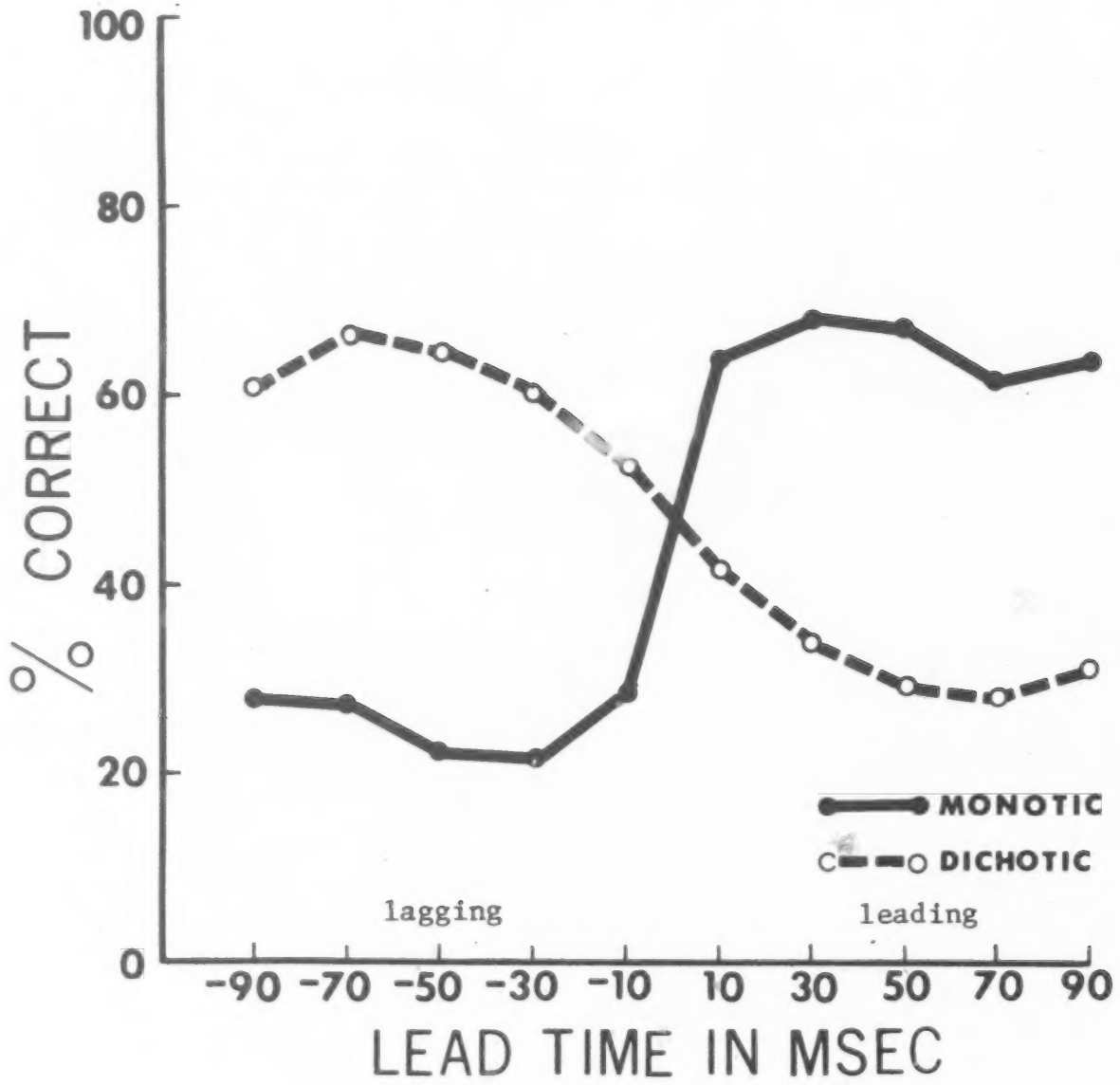


Fig. 4

The monotic lead effect and dichotic lag effect seen in Figure 3 are enhanced in the clarity judgments. It will be seen in Experiments 3 and 4 that the pattern of clarity judgments is a much more sensitive measure of the lag effect than the overall percent correct. For this reason much of the subsequent analysis is based entirely on first responses.

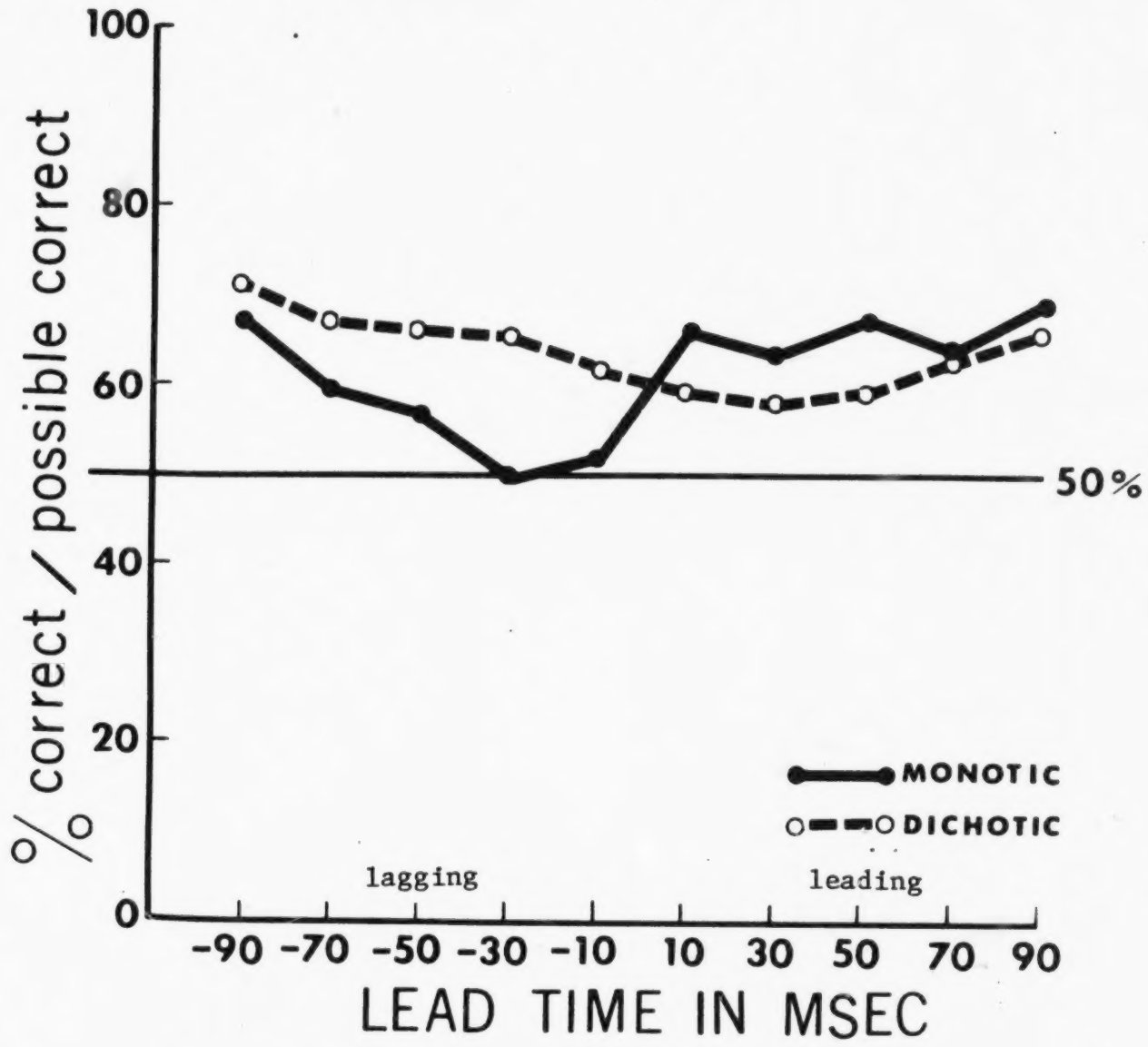
Analysis of second responses. The subjects were instructed to record two different responses on each trial even if they had to guess. According to the instructions, all guesses should have been recorded in the second column of the answer sheet. Thus, of the responses in the second column, many were undoubtedly simply guesses, although others represent stimuli which were perceived correctly but judged as less clear than the first responses. In order to separate the guesses from the correct second responses, it is necessary to consider the pattern of first responses. The frequency with which a lagging or leading stimulus would be correctly guessed as a second response depends on the frequency of lagging and leading stimuli given as first responses. For example, on the dichotic test, lagging stimuli were given as first responses 61.1 percent of the time, so that lagging stimuli could be given as a second response on not more than 38.9 percent of all second responses. Leading stimuli were given as the first response on only 33.3 percent of the trials and thus could potentially be given as the second response on 66.7 percent of the trials. If all second responses were guesses, second responses would contain 33.3 percent leading stimuli, 19.5 percent lagging stimuli, and 47.5 percent errors. Thus, simply by chance, the lag advantage would appear to be reversed in second responses.

An analysis of second responses was performed which takes into account the fact that there are an unequal number of trials in which lagging and leading stimuli could appear as the second responses, given the first response pattern. Accuracy of identification for second responses was calculated as number of $\frac{\text{correct}}{\text{possible correct}}$ second responses. A rise above 50 percent shows that stimuli which were judged as less clear were reported with better than chance accuracy.

Figure 5 shows the percent correct second responses for the dichotic and monotic tests after the scores have been corrected in the manner just described. The first response patterns--the monotic lead effect and the dichotic lag effect--also appear in the second responses. On the monotic test, stops which were lagging by 10 or 30 msec were reported with no more than chance accuracy as second responses, but stops leading by all intervals were given as second responses much more often than would be expected if subjects were guessing. On the dichotic test it would appear that subjects more often guessed about the second response when the leading stop would have been correct than when the lagging stop would have been correct.

Individual differences in performance on the monotic and dichotic tests. Up to this point the comparison of the dichotic and monotic tests has considered the averaged performance of all subjects. It is also of interest to look at the distribution of lag effects and lead effects among individuals within each of the test groups. A lag effect or lead effect score was computed for each subject using as the measure of these effects the expression: (Leading-Lagging)/(Leading + Lagging) where "Lagging" and "Leading" refer to the number of first responses corresponding to lagging or leading stops summed over all delay intervals. The distributions of these scores for the dichotic

Mean Percent Correct Second Responses for the Dichotic and Monotic Tests



Note: The percent correct at each onset time has been corrected for the pattern of first responses by using the formula percent (correct/possible correct) second responses. With this correction the probability of giving a correct second response by chance is .50 at each onset time condition.

Fig. 5

and monotic tests are shown in Figure 6. These distributions give further evidence of the reliability of the monotic lead effect and the dichotic lag effect. All twelve of the subjects on the monotic test had lead effects. On the dichotic test, eleven of the twelve subjects had lag effects; one subject had a lead effect on the dichotic test, but this was smaller than any of the monotic scores. There was also a difference in variability between the two conditions, with greater variability among subjects in the dichotic than in the monotic condition.

Distribution of dichotic laterality effects. In previous dichotic listening studies, report of stop consonants at the right ear was found to be more accurate than report of stops at the left ear. Evidence of ear asymmetry was, therefore, sought in the present experiment. The number of subjects showing a right-ear advantage was established by computing a laterality effect score for each subject. The size of the ear effect was calculated from the formula (Right - Left)/(Right + Left) where "Right" and "Left" refer to the number of first responses corresponding to stimuli delivered to the right and left ears.[4] The index ranges from .00, for no difference between ears, to $\pm$ 1.00, which would indicate that all first responses are from the right ear (positive scores) or left ear (negative scores). Figure 7 gives the frequency distribution of these laterality effect scores. Of the twelve subjects, ten had a right-ear advantage and two a left-ear advantage. For the subjects with left-ear effects, the extent of ear asymmetry was less than that seen for the typical right-ear effect.

Relationship between the ear effect and lag effect. The fact that both a right-ear effect and lag effect are reliably observed with dichotic presentation of temporally staggered stop consonants leads one to inquire into the nature of the relationship between the ear effect and lag effect. One possibility is that the two phenomena are essentially independent. On the other hand, interaction between the two effects might be observed.

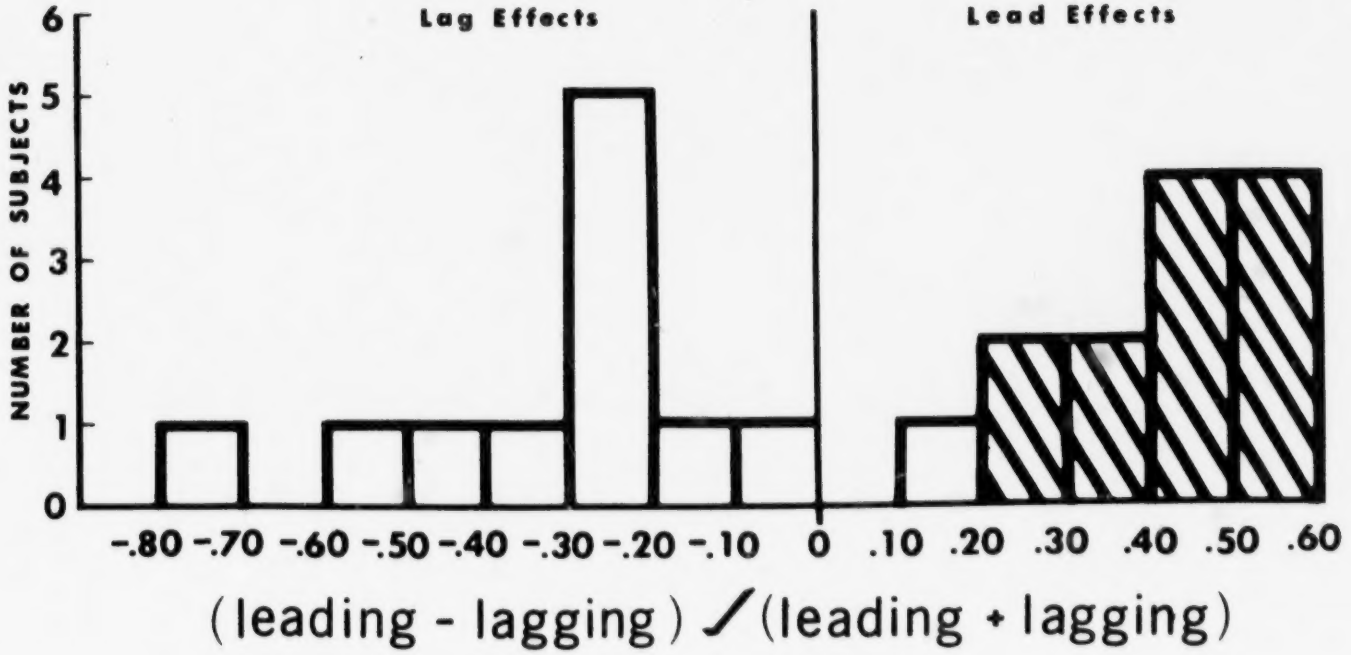
One way of approaching this question was to determine whether the subjects who had large lag effects were also the subjects who had large ear effects. A Spearman rank correlation coefficient (Siegel, 1956) was computed to determine the extent of association between the lag effect and right-ear effect scores. The coefficient was -.26, a nonsignificant correlation. The negative sign probably reflects a ceiling effect imposed by the task; extremely large lag effects and ear effects are mutually exclusive. Thus, the subjects with the largest ear effects had small lag effects and vice versa. However, subjects with moderate ear effects and lag effects did not show any correlation in the size of the two effects.

Effects of relative onset time on the perception of syllables at the right and left ears. Another way of looking for an interaction between the ear effect and lag effect is to see whether the consequences of interaural delay are the same for the two ears. Figure 8 shows the percentage of first responses from the right and left ear for each delay condition for all twelve

---

[4]This measure of the right-ear effect was introduced by Studdert-Kennedy and Shankweiler (1970).

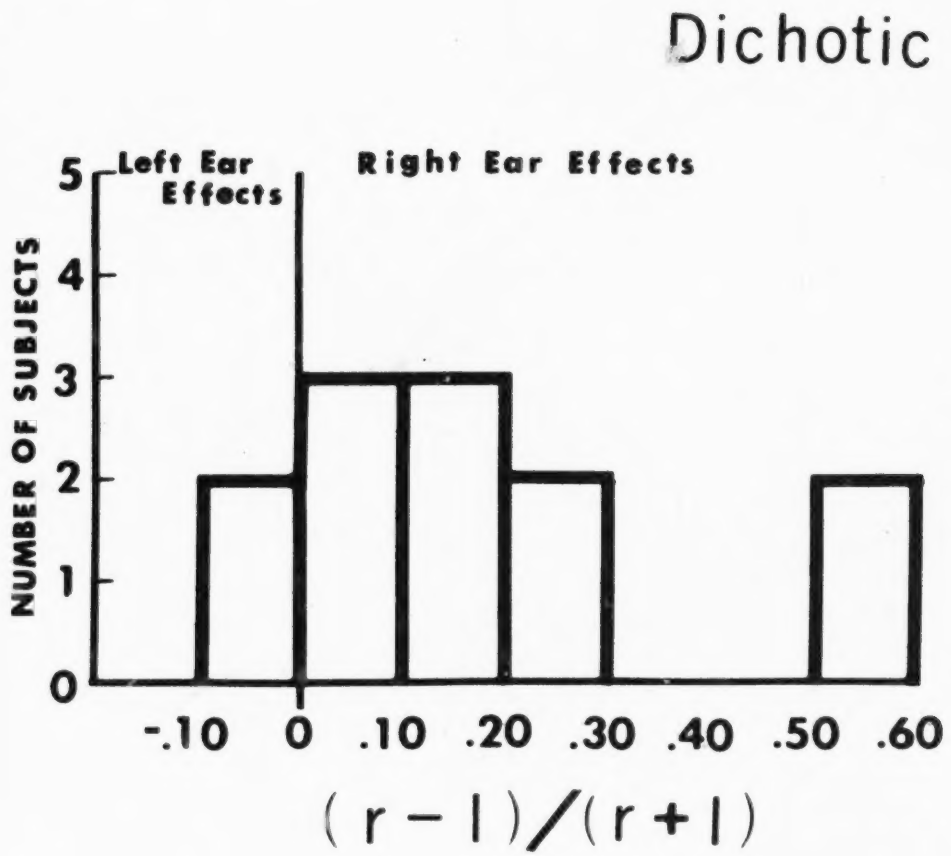
Frequency Distributions of Lag Effect Scores, (Leading-Lagging)/(Leading+Lagging), for the Dichotic and Monotic Conditions



Note: Positive scores indicate lead effects and negative scores indicate lag effects. Scores are based on first responses.

Fig. 6

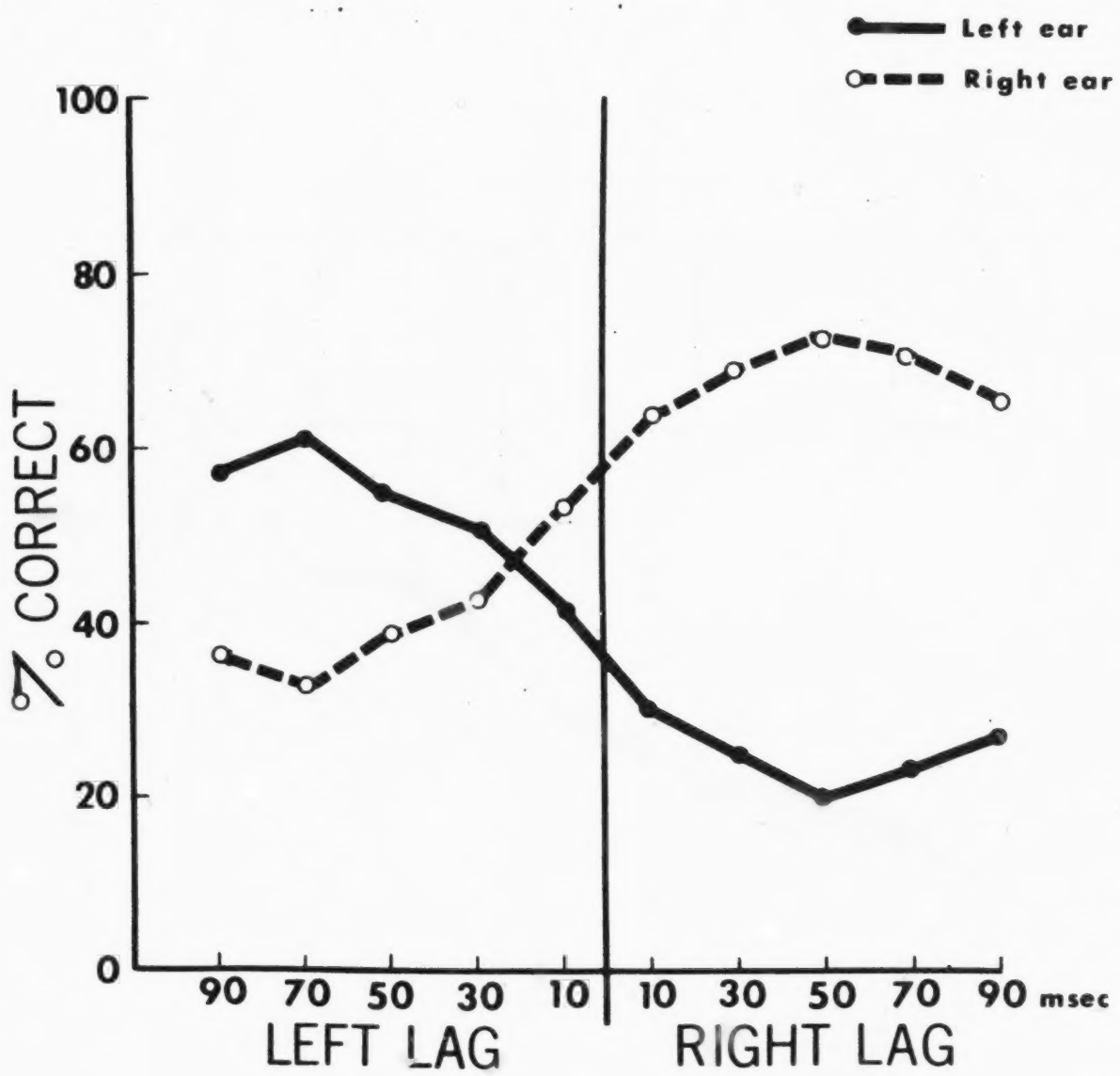
Frequency Distribution of Ear Effect Scores, (R-L)/(R+L), Based on First Responses in the Dichotic Condition



Note: A positive score indicates a right-ear effect and a negative score indicates a left-ear effect.

Fig. 7

Comparison of Left-Ear Lag and Right-Ear Lag Trials in the Dichotic Condition



Note: The figure shows the mean percentage of first responses corresponding to stimuli at the left and right ears at each delay interval.

Fig. 8

subjects on the dichotic test. It can be observed that stimuli at the right ear were given as first responses more often than stimuli at the left ear but that a left-ear lag can counteract the right-ear effect. Thus, with a 10-msec separation between ears, stops at the right ear were judged to be clearer than those at the left, but when the left ear was delayed 30 msec behind the right, stimuli at the left ear were judged as clearer.

Figures 9 and 10 show the right and left ear curves for two of the subjects. One of the subjects (SV) performed in a manner very similar to the group performance, but the other subject (BR) was remarkable in the magnitude of the right-ear effect. BR judged the right ear to be clearer than the left even when the left ear was lagging behind the right. Nevertheless, BR showed some improvement for the left ear when it received the lagging syllable. Thus, both subjects showed a right-ear advantage and an advantage for the lagging ear, regardless of which was the delayed ear.

Figure 11 plots the data in Figure 8 in a different format which compares the frequency of report of the two ears under identical delay conditions. For any delay between ears the right ear is superior to the left, but the shape the function relating frequency of first response to relative onset time is similar for the two ears. There is one noticeable difference between the ears--the consistent 20-msec displacement between the left- and right-ear curves. The maximum right-ear advantage occurs with a right-ear delay of 50 msec, while the maximum left-ear advantage occurs with a left-ear delay of 70 msec. Figure 12 compares the curves for the two ears after all left-ear points have been advanced 20 msec along the x-axis. Following this displacement it is observed that the curves for the two ears have the identical form.

Discussion

The results of Experiment 1 confirm the earlier finding that when two stop consonant-vowel syllables sharing the same vowel are overlapped in time, a lag effect is obtained with dichotic presentation and a lead effect is obtained with monotic presentation. These contrasting effects of monotic and dichotic competition were observed in the overall frequency of correct responses, in judgments of the relative clarity of the two stops, and in the accuracy of identification of the "less clear" stimulus.

Other differences between the dichotic and monotic conditions were discovered in Experiment 1. With dichotic presentation the maximum advantage for the lagging syllable occurred with delays of 50 to 70 msec between syllable onsets, whereas the monotic lead effect was maximal with shorter delays (10 to 30 msec). At these short delay intervals the lagging stop was completely obscured on the monotic test, as indicated by the fact that the lagging stimuli were reported correctly no more often than would be expected by chance. In contrast, the leading syllables on the dichotic test were reported with greater than chance accuracy at all delays. Finally, it was found that the performance of individual subjects with respect to the magnitude of the lag or lead effect was more variable with dichotic presentation than with monotic presentation.

It is not clear why there should be such a great difference between the effects of dichotic and monotic competition. Studdert-Kennedy et al. (1970)

Percent Correct First Responses by Ear on the Dichotic Test for Subject SV

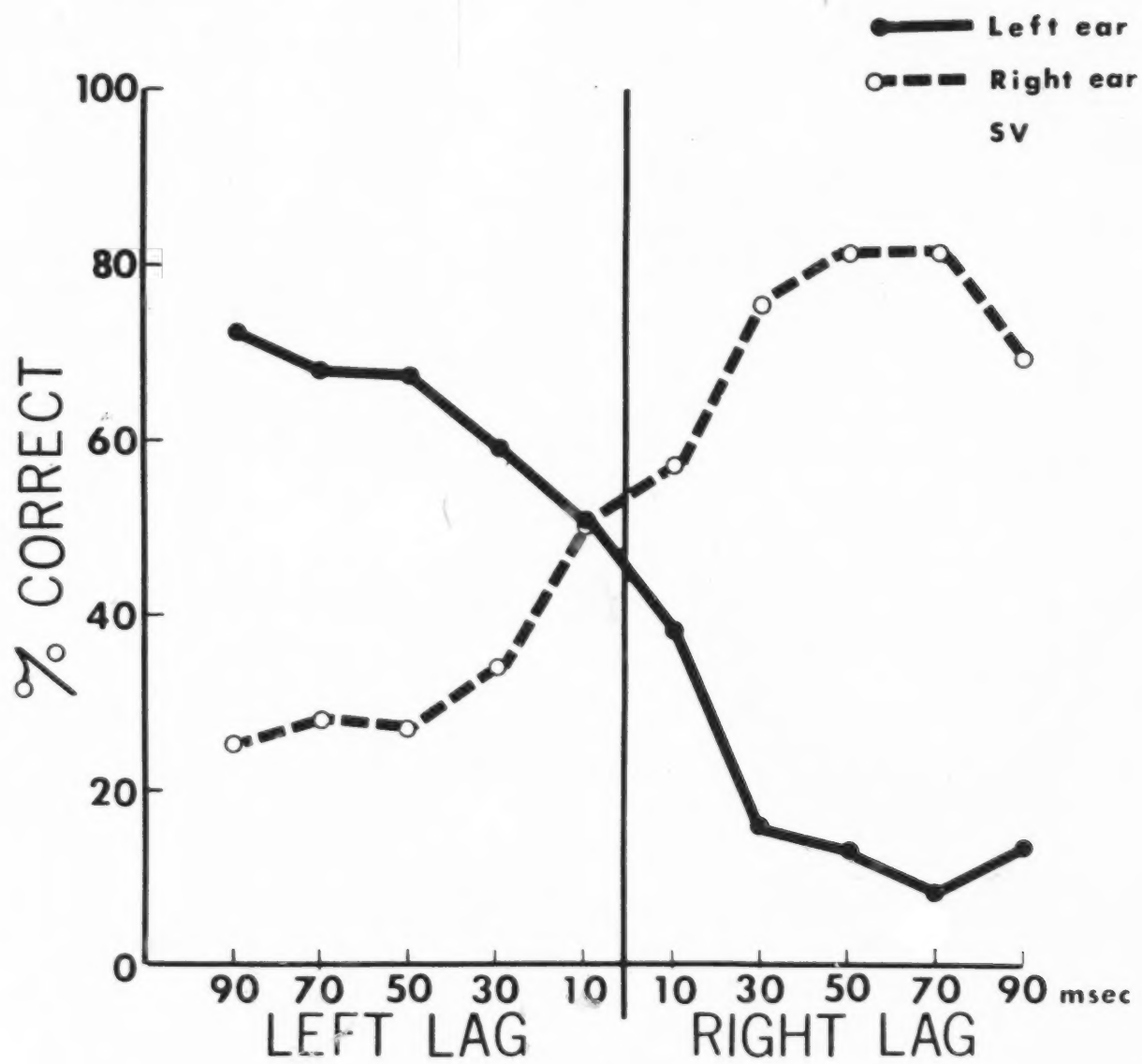


Fig. 9

Percent Correct First Responses by Ear on the Dichotic Test for Subject BR

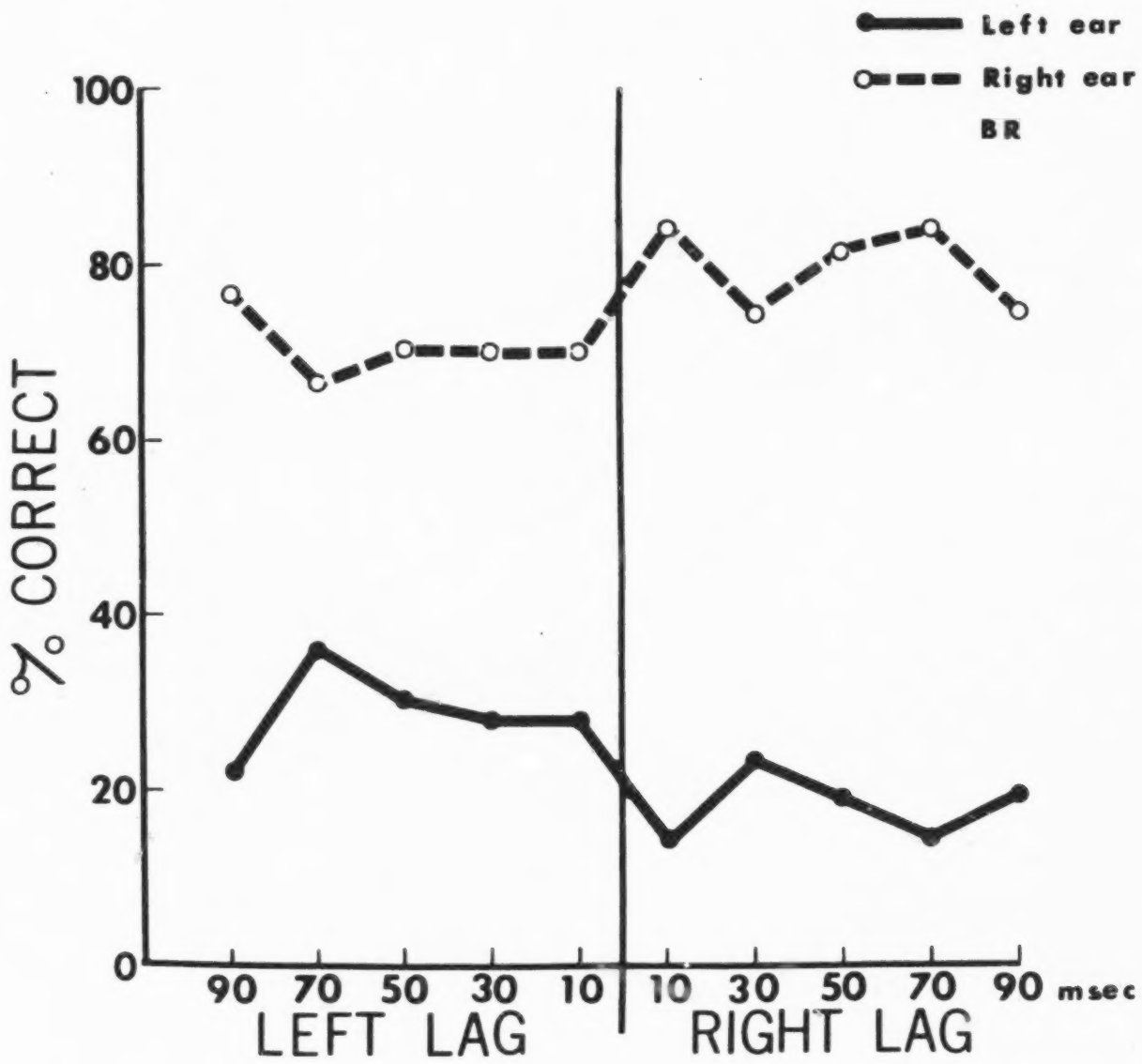


Fig. 10

Mean Percent Correct First Responses by Ear Comparing the Same Delay Conditions for the Two Ears

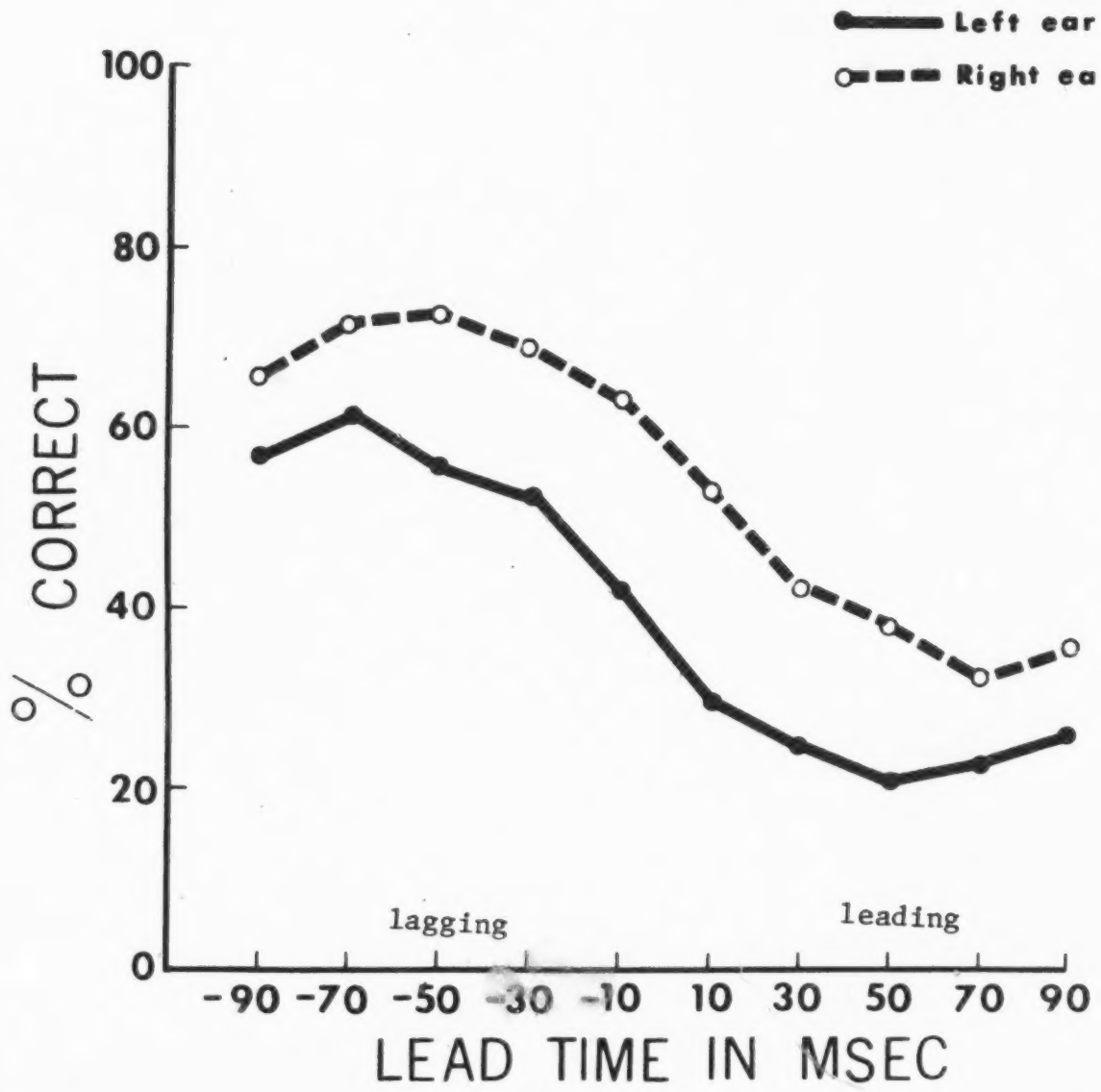


Fig. 11

Comparison of the Form of the Lag Effect Function for the Two Ears After a Displacement of the Left-Ear Curve 20 msec Along the X-Axis

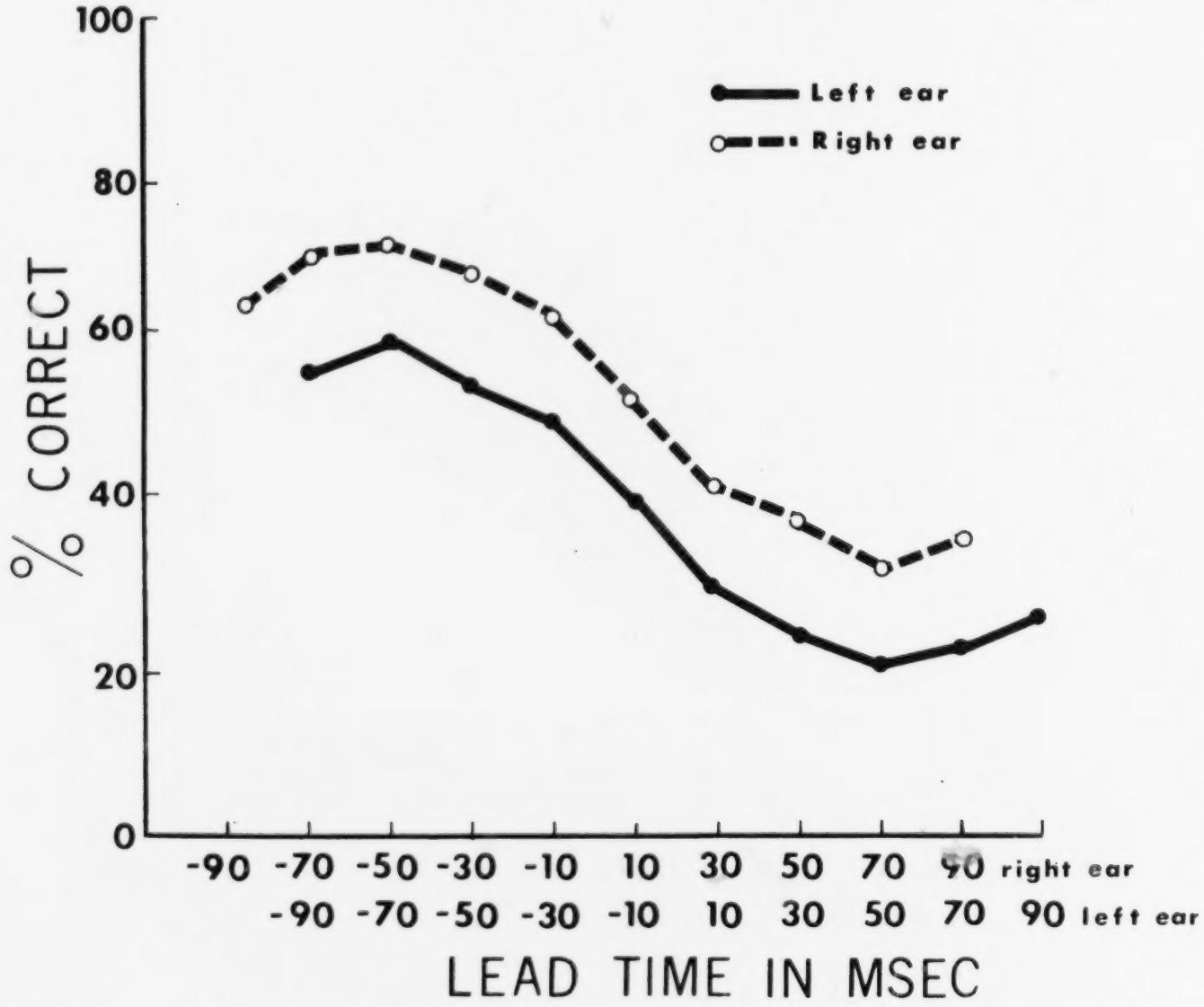


Fig. 12

have attributed the monotic lead effect to peripheral masking and the dichotic lag effect to central processes. It is evident that the lag effect must arise centrally because with dichotic presentation the stimuli do not interact at the peripheral receptor. The puzzling question is why there is no evidence of a central lag effect with monotic competition as well. Perhaps the lag effect occurs only when the stimuli arrive over separate input channels, or the explanation may be that the peripheral suppression of the lagging stops is so extensive with monotic presentation as to preclude subsequent central enhancement of the lagging stimuli.

The finding of a right-ear advantage with dichotic presentation was also consistent with the results of previous research. The stops at the right ear were judged as clearer than those at the left ear for all delays when the same delay conditions were compared for the two ears. In comparing the effects of delay time for the two ears it was found that the shape of the function relating delay time to the percent correct first responses was the same for the left and right ears. From this result one might infer that the lag effect and right-ear effect are separate phenomena which combine in an additive manner to determine the relative intelligibility of the competing stop consonants. Further evidence of the independence of the two effects was the finding that individual differences in the lag effect and right-ear effect were not correlated. The only data suggesting an interaction between the lag effect and right-ear effect were the results dealing with the location of the maximum lag effect for the two ears. The maximum lag effect for the left ear was at a left-ear lag of 70 msec, while the maximum lag advantage for the right ear was with a right-ear lag of 50 msec. Certain findings in Experiment 2 lead to the conclusion that this temporal displacement is an artifact of the particular delay intervals tested and that the true peak is in the same place for the two ears.

EXPERIMENT 2. SELECTIVE LISTENING FOR DICHOTICALLY PRESENTED STOP CONSONANTS

Experiment 1 showed that when CV syllables are presented dichotically with slight separations in onset time between ears, lagging consonants are judged as clearer and are more accurately identified than leading consonants. It is clear that this lag effect in report of dichotically presented stop consonants is a central phenomenon, but the precise level of processing at which the lag effect originates has not yet been established. On the one hand, it seems reasonable to suppose that the lag effect reflects a perceptual advantage for the lagging syllable. On the other hand, one might also suppose that the lag effect arises in the organization of responses rather than in the course of analysis of the stimuli. If the latter interpretation were correct, then one would expect to obtain a lag effect only when the subjects are required to report both stimuli on each trial, as was the case in Experiment 1. The lag effect could be explained by supposing that the subjects have a bias toward reporting the lagging stimulus first and the leading stimulus second. As a consequence of this order of report, leading stimuli would be more often forgotten than lagging stimuli.

The view that the lag effect results from a bias in ordering of responses is not consistent with some of the findings in Experiment 1. It was found, for example, that the advantage for lagging syllables decreased with delays greater than 50 msec; if a bias favoring lagging syllables were the sole explanation of the lag effect, the effect should have increased still further with longer delays. Furthermore, evidence of a lag advantage was found in second responses as well as first responses.

Although it would seem more plausible to attribute the lag effect to features of the perceptual processing rather than to response organization, nevertheless, the method used in Experiment 1 leaves some ambiguity in this regard. Experiment 2 repeated the dichotic listening test but changed the method of report so that only one response was required on each trial. This was accomplished by transforming the experiment into a selective listening task. Experiment 2 contained two test conditions. In one condition the subjects were asked to listen exclusively for the stop consonants arriving at one ear and to ignore those at the other ear. In the other condition the subjects were instructed to attend to the order of arrival of the two stops within each pair and report either the leading stops or the lagging stops according to instructions.

A comparison of the performance on the selective listening tasks with the results of the clarity judgment procedure provides an opportunity to assess the generality of the lag effect with rather dissimilar recall methods. Moreover, if the subjects proved to be more accurate in selective report of lagging syllables than leading syllables, this would give stronger evidence of a perceptual basis of the lag effect.

Method

Each subject participated in four one-hour test sessions, two sessions for the temporal order task and two for the ear-monitoring task. The stimulus tape used in these tests was the same as that used in Experiment 1. Each session consisted of two runs through the 180-trial tape with headphones reversed on the second run, giving a total of 360 trials for each session. Each session was split into four blocks of 90 trials. Instructions were given to the subjects at the beginning of each block of 90 trials.

Ear-monitoring task. On the ear-monitoring task the subjects were told that they would receive different stop consonants at each ear on every trial but that they were to report only those arriving at a designated ear. They were informed at the beginning of each block of ninety trials which ear they were to report. The blocks were arranged in the order R-L-L-R or L-R-R-L, where L and R refer to instructions to report the left or right ear. The subjects were randomly assigned to one of these arrangements for the first ear-monitoring session and were automatically assigned to the other arrangement in the second session. Within any session the two ears had exactly the same combination of stimulus conditions and recall instructions, but two sessions were necessary to present all combinations of syllable pairs, delays, and instructions.

The subjects were required to give one response on each trial even if they had to guess and to respond only with "B," "D," or "G."

Temporal order task. On the temporal order task the subjects were told that they would receive two different syllables on each trial and that the syllables would differ between ears in time of arrival. For two blocks of ninety trials within each session they were to report only the leading stop consonant, and on the other two blocks of trials they were to report only the lagging stop. The blocks were arranged in the order 1-2-2-1 or 2-1-1-2, where 1 and 2 indicate instructions to report the first or second stop consonant. Subjects were randomly assigned to one of these arrangements on the first session of the temporal order task and to the other arrangement on the second session.

Here again the subjects were required to give one response on each trial even if they had to guess and to respond only with "B," "D," or "G."

Subjects. The same subjects took both the temporal order and ear-monitoring tasks, half taking the ear-monitoring task first and the other half taking the temporal order task first. There were twelve subjects, all of whom were right handed. The subjects were introductory psychology students at the University of Connecticut.

Results

Accuracy of selection. On each trial two different stops were presented, but only one of those two was correct according to the selection instructions. If the subject reported the stimulus which he should have ignored according to

the instructions, his response on that trial is termed an intrusion. The most frequent errors on the selective listening tasks were intrusions (failures to judge correctly the ear or order of arrival). Intrusions occurred on 38 percent of the temporal order trials and 32 percent of the ear-monitoring trials. Responses which were correct according to the selection instructions were given on 52 percent of the temporal order trials and on 62 percent of the ear-monitoring trials. Responses corresponding to neither of the stimulus consonants were given on fewer than 10 percent of all trials.

An excess of correct responses over intrusions indicates that the subject was selecting the response in accordance with the instructions. Figure 13 shows the extent to which correct responses exceeded intrusions as a function of the interval between syllable onsets. Accuracy of selection improved consistently with increasing delays between syllables. From the increase in accuracy with longer delays, it can be inferred that the subjects were attempting to perform in accordance with the selection instructions. However, it is evident from the low accuracy scores that they experienced considerable difficulty in selecting correctly, particularly when syllable onsets were close in time. Selection by temporal order was more difficult than selection by ear for ten of the twelve subjects.

Effect of the delay between syllables and ear of arrival of the selected stimulus on the accuracy of selection. Figures 14 and 15 show the pattern of results obtained under each of the selection instructions. Figure 14 shows that the response was more often correct when subjects were attending to the right ear than when they were attending to the left ear. Moreover, report from either ear was more accurate when that ear was the one receiving the lagging syllable. Figure 15 indicates that the subjects were more often correct when instructed to report the lagging syllable than when instructed to report the leading syllable. Reporting either the lagging or leading syllable was easier when that syllable was arriving at the right ear than when it was arriving at the left ear.

Comparison of the ear-monitoring and temporal order tasks. Figures 14 and 15 give clear evidence that the accuracy of selection is dependent upon the relative onset time and ear of arrival of the selected stimulus. The present section compares the two tasks with respect to the influence of the lag effect and ear effect. Figure 16 shows how often subjects gave the correct response as a function of ear and time of arrival for the ear-monitoring and temporal order tasks. Correct responses were more frequent for ear-monitoring than for temporal order judgments, but the variations in accuracy of selection attributable to the ear of arrival and onset time of the selected stimulus are remarkably similar for the two tasks. This similarity between tasks is also apparent in Figure 17 which shows the distribution of intrusions according to the ear and onset time of the intruding syllable.

In order to assess more precisely the influence of the lag effect and ear effect in these tasks, the data were tabulated again in such a way as to separate the lag effect and ear effect from each other. Each response was first categorized as corresponding to a stimulus at the left ear or right ear or as an error--regardless of whether it was a correct response or an intrusion and regardless of whether it corresponded to a leading or lagging stimulus. The difference between the frequency of responses from the two ears at each delay interval was taken as the measure of the ear asymmetry at that temporal offset. Likewise, the lag effect at each

Accuracy of Selecting Responses by Ear and by Temporal Order as a Function of the Time Between Syllable Onsets

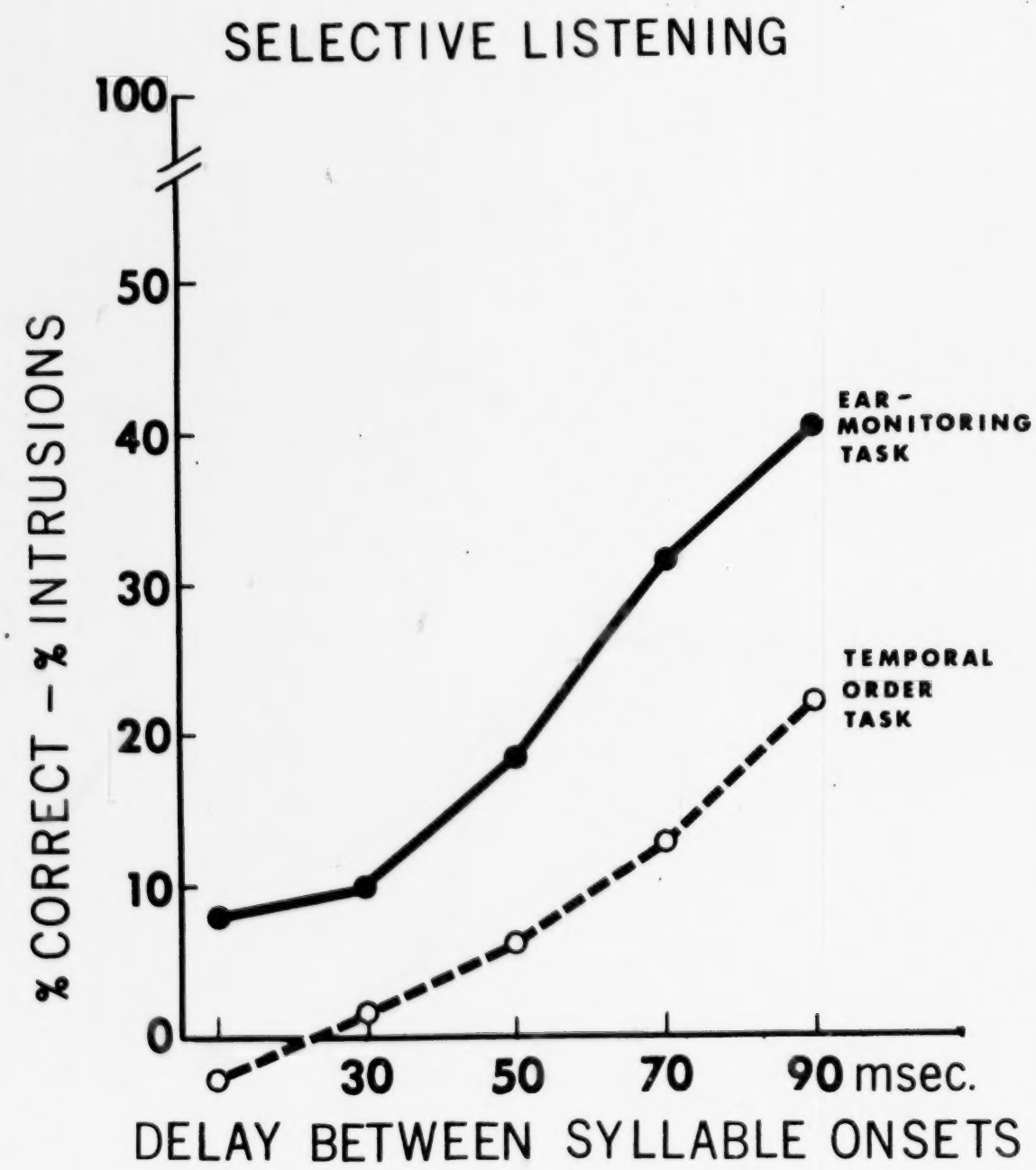
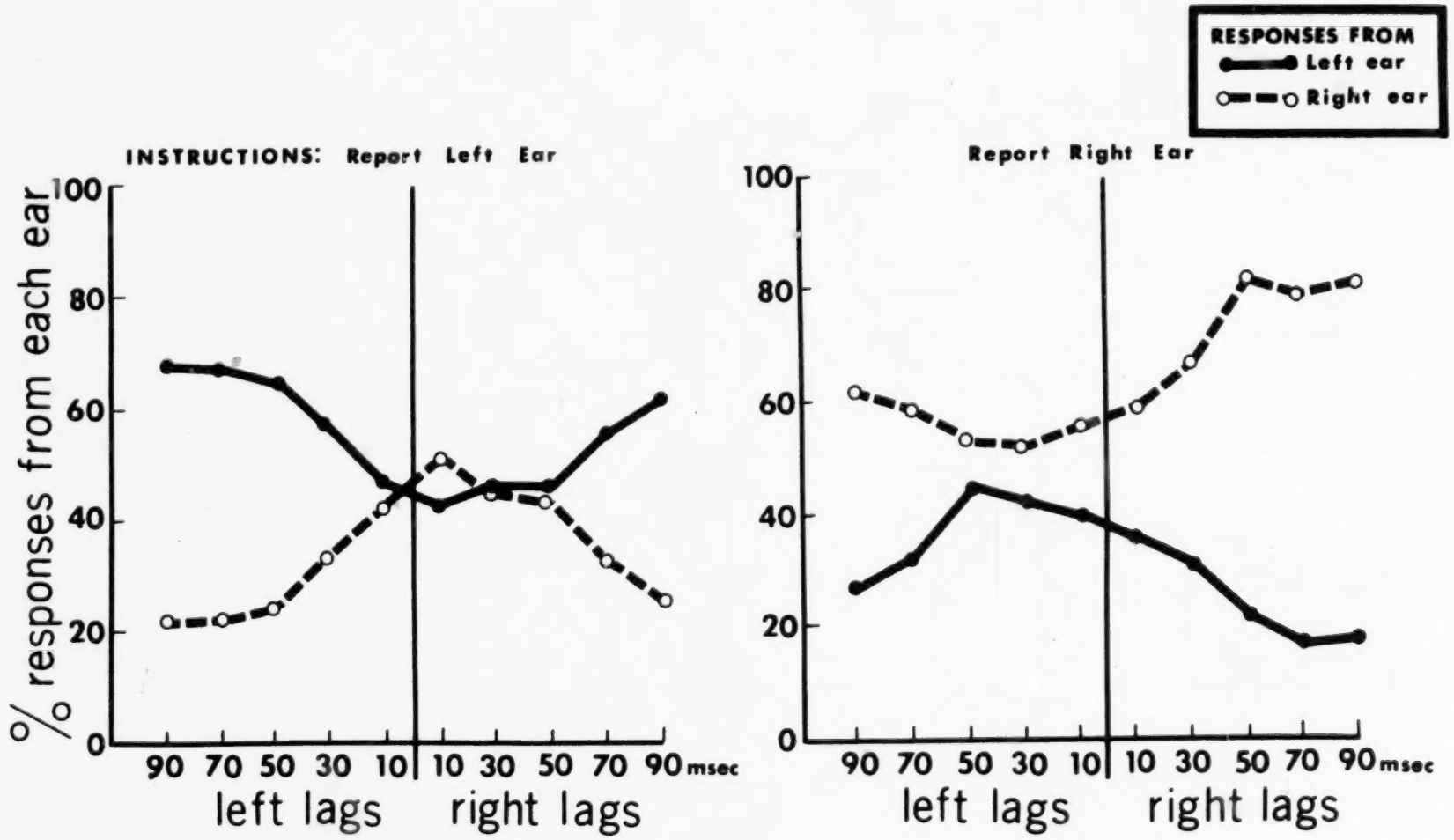


Fig. 13

Accuracy of Selecting Responses from the Left and Right Ears



Note: The figure shows the relative frequency of responses from the left and right ears depending on which was the attended ear, which ear was lagging, and the amount of time between syllables.

Fig. 14

Accuracy of Selecting Lagging and Leading Stimuli

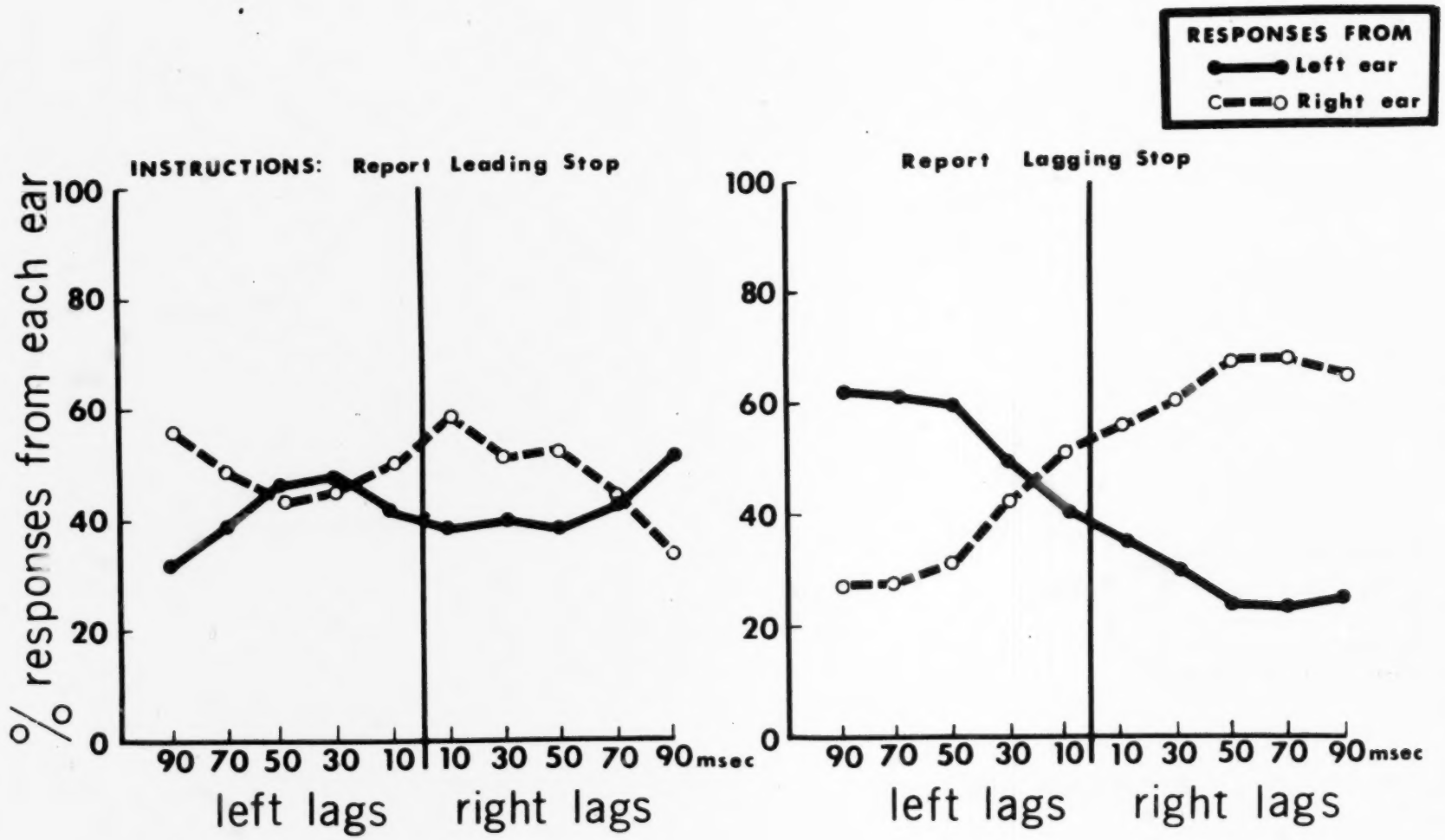


Fig. 15

Note: The figure shows the relative frequency of responses from the left and right ears depending on which ear received the lagging syllable, whether instructions were to report the lagging or leading syllable, and the amount of time between syllable onsets.

Mean Percent Correct Responses on the Ear-Monitoring and Temporal Order Tasks as a Function of Ear and Relative Onset Time of the Correct Syllable

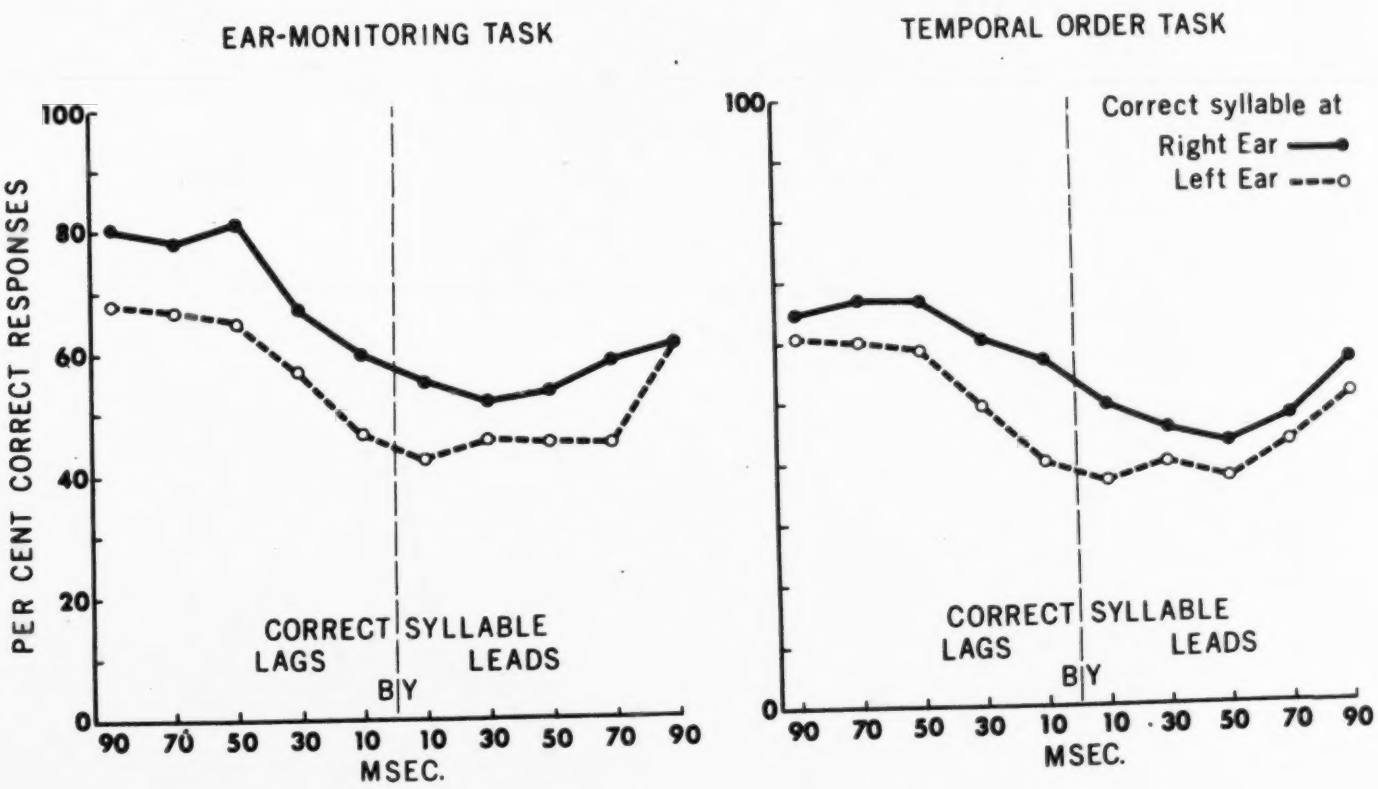


Fig. 16

Mean Percent Intrusions on the Ear-Monitoring and Temporal Tasks as a Function of the Ear and Relative Onset Time of the Intruding Syllable

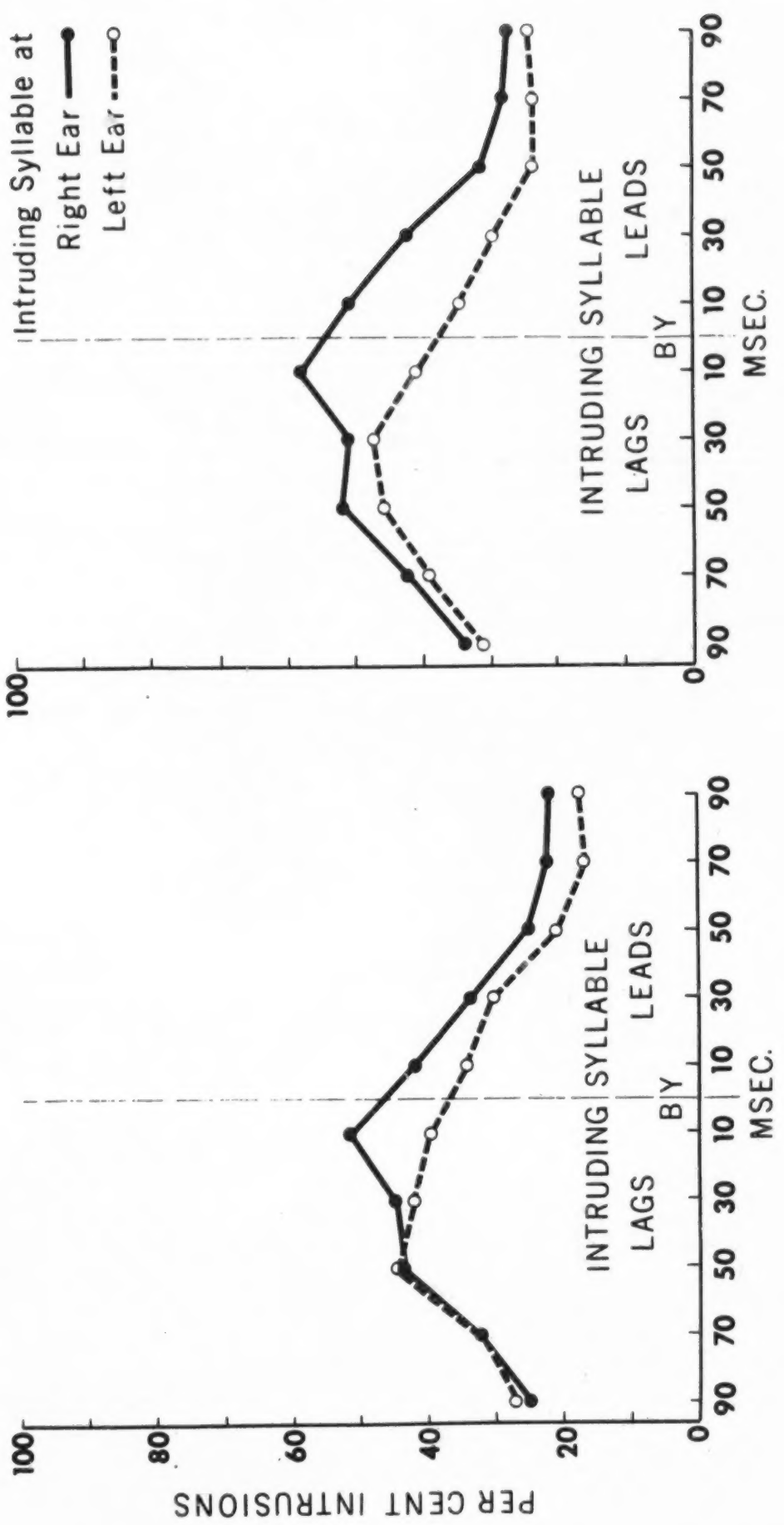


Fig. 17

delay was measured by categorizing each response as corresponding to a lagging stimulus or a leading stimulus and computing the difference between the number of lagging and leading stimuli given as responses. This analysis was carried out only for the seven subjects who had right-ear effects in both the ear-monitoring and temporal order tasks.

Figure 18 shows two graphs, one for the ear effect and one for the lag effect. A striking feature of these graphs is the similarity between the two tasks. Also noteworthy is the fact that the time course differed greatly between the lag effect and ear effect. The right-ear advantage was greatest when syllables were most nearly simultaneous (10-msec delay) and decreased monotonically with longer delays. The lag effect increased with delays up to 50 msec and declined with still longer delays. The variations in the lag and ear effects as a function of delay interval were similar for the ear-monitoring and temporal order tasks.

Individual differences on the ear-monitoring and temporal order tasks. The behavior of each of the twelve subjects was compared across the two tasks on three measures of performance--accuracy of selection, ear effect, and lag effect. An index of each measure was computed for each subject on each task as in Experiment 1. The index of accuracy of monitoring was calculated using the formula (Correct - Intrusions)/(Correct + Intrusions), where the words "Correct" and "Intrusions" refer to the number of responses falling into each of those two categories. Similarly, the lag effect index was computed as (Leading - Lagging)/(Leading + Lagging) and the ear effect index as (Right - Left)/(Right + Left).

On all three measures, the performance of the individual subjects was extremely consistent across the two tasks. A Spearman rank correlation coefficient was computed for each of the three measures to assess the consistency in ranking of the individual subjects on the two tasks. The Spearman rank correlation coefficient relating the accuracy of selection by ear and by temporal order was +.79 ($p<.01$). This means that the subjects who were accurate in selecting by ear also tended to be those who were accurate in selecting by order of arrival. The Spearman rank correlation coefficient for the right-ear effect indices across the two tasks was +.97 ($p<.001$). For the lag effect the correlation between the two tasks was +.65 ($p<.05$).

Comparison of the selective listening and clarity judgment methods. It might be expected that the magnitude of the lag effect and ear effect would depend on the method of report. We have seen that there was no difference between the magnitude of these effects in the two selective listening methods--the ear-monitoring and temporal order tasks. However, differences were found between the clarity judgment method and selective listening.

A comparison was made of the frequency distributions of lag effect scores obtained with the selective listening and clarity judgment procedures. Since the ear-monitoring and temporal order tasks gave such similar results, only the ear-monitoring data were considered in this comparison and ten more subjects, in addition to the twelve already described, were run on the ear-monitoring task. This gave a total of twenty-two subjects on the ear-monitoring task. These were compared with twenty-four subjects on the clarity judgment task (the twelve subjects from Experiment 1 and twelve more from Experiment 3). Figure 19 shows the frequency distributions of lag

Changes in the Magnitude of the Right-Ear Effect and Lag Effect as a Function of Interaural Delay Time for the Ear-Monitoring and Temporal Order Tasks

Means for Seven Subjects with Right-Ear Effects on Both Tasks

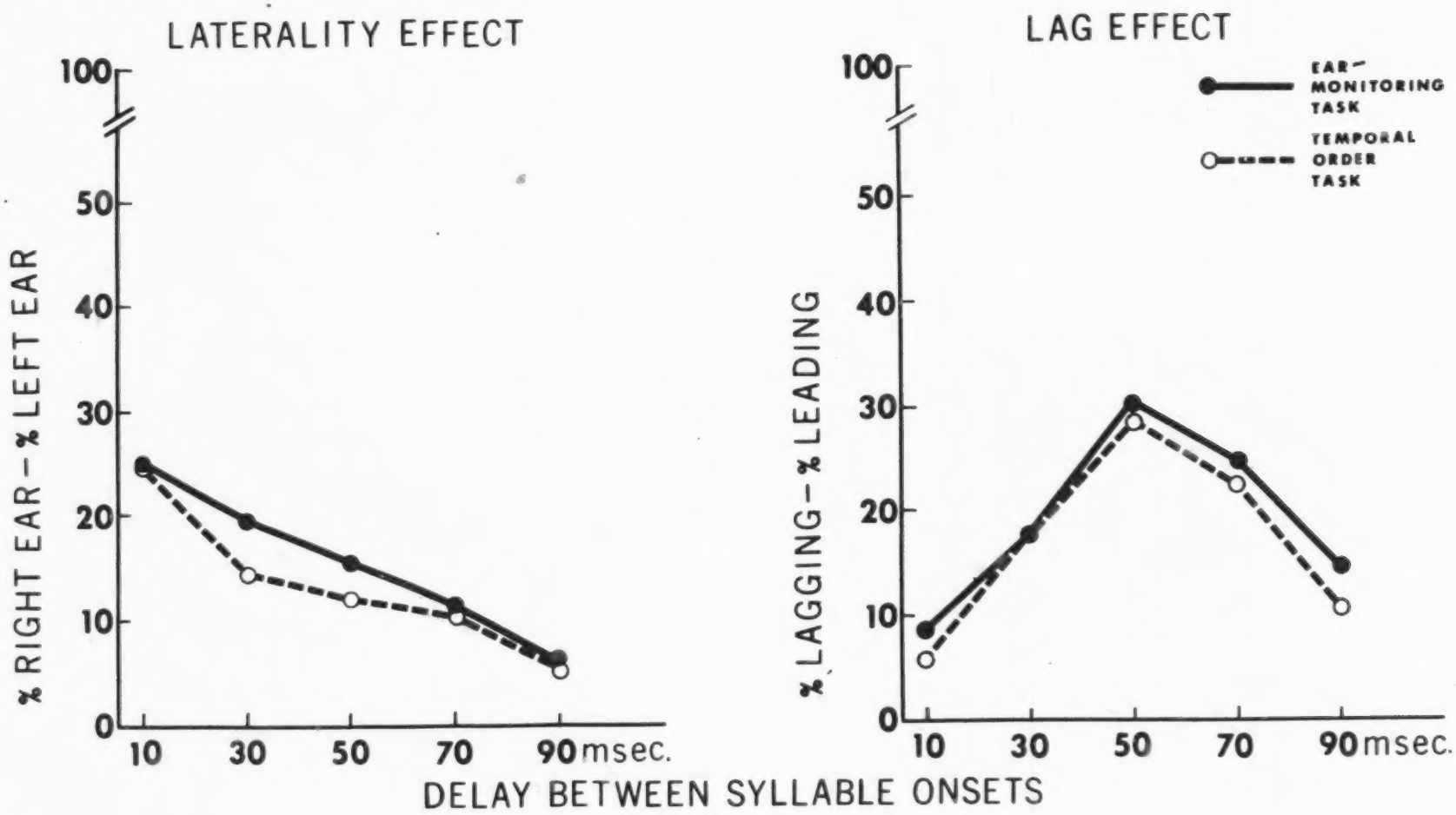
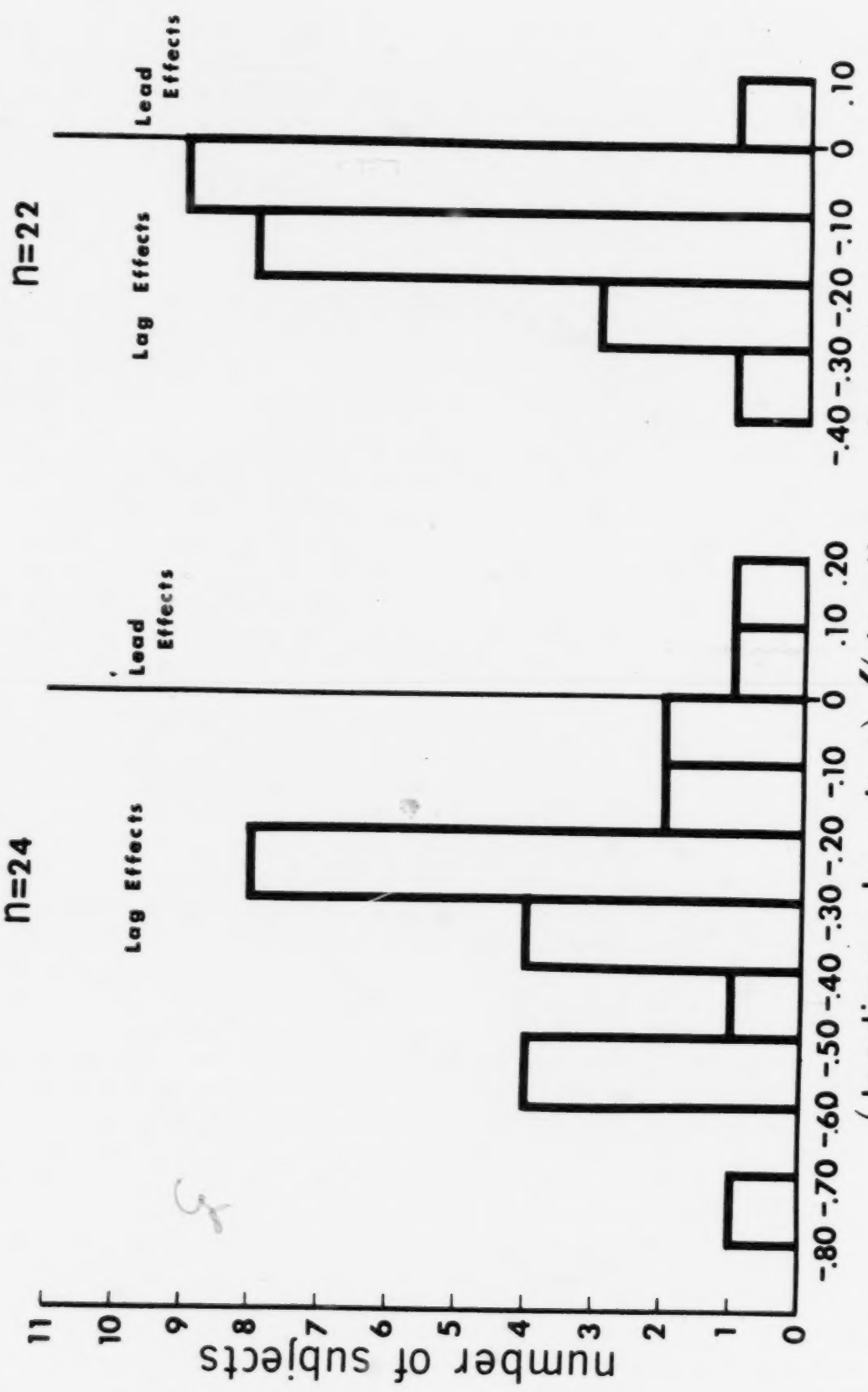


Fig. 18

Comparison of the Frequency Distributions of Lag Effect Scores Obtained with Ear Monitoring and Clarity Judgments



Note: Negative scores indicate lag effects and positive scores indicate lead effects.

Fig. 19

effect scores for the ear-monitoring and clarity judgment tasks. The lag effect scores on the ear monitoring task were significantly smaller by a Mann-Whitney U test (Siegel, 1956) than the lag effects obtained with clarity judgments ($z=3.38$, $p < .001$, two-tailed).

A comparison was also made of the distributions of ear effect scores for the two tasks (Fig. 20). Although the right-ear effects were somewhat smaller with the ear-monitoring method, the difference between ear effect scores for the two methods was not statistically significant ($z = 1.54$, $p < .124$, two-tailed).

Discussion

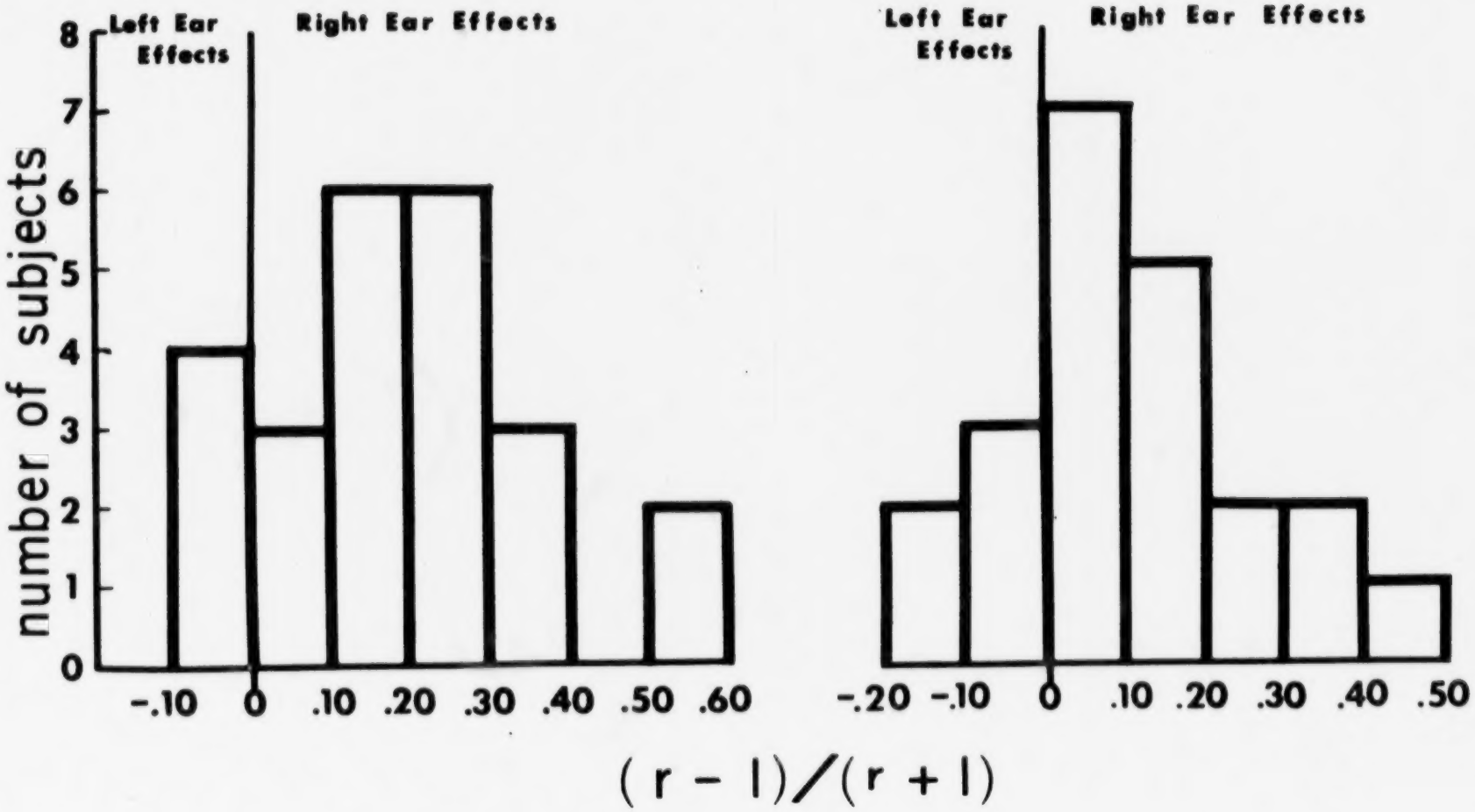
Experiment 2 demonstrates that the lag effect does not depend upon any particular method of report. The lag effect was found in two selective listening tasks which were quite different from each other and which also differed greatly from the clarity judgment task described in Experiment 1.

The subjects exhibited a sensitivity to relative onset time with delays as short as 10 msec between ears as indicated by the more frequent report of lagging than leading syllables at that delay. However, the subjects were unable to judge which was the leading or lagging syllable with delays as long as 50 msec between ears, and even with 90-msec delays, judgments of temporal order were inaccurate. It is clear, therefore, that the lag effect does not depend on perception of temporal order. Furthermore, a lag effect was obtained when attention was directed toward ear of arrival, a dimension unrelated to temporal order.

By the same reasoning it is evident that a response bias or attentional bias cannot account for the right-ear effect, although Inglis (1965) did offer such an explanation of laterality effect. In Experiment 2, the right-ear advantage was found to be greatest at short delay intervals, where confusions between ears were most common. Thus, the intervals which gave the largest right-ear effects are those intervals at which the subjects would have found it most difficult to bias their responses in favor of the right ear. These results point clearly to an interpretation of both the lag effect and the right-ear effect as perceptual phenomena.

One striking finding in this experiment was the difficulty which subjects experienced in performing the selective listening tasks. It is perhaps not surprising that the task of reporting by order of arrival should prove difficult. Hirsch (1959) studied temporal order judgments for simple nonspeech stimuli, such as pure tones, and found that at least 20 msec must elapse between stimulus onsets in order to obtain correct temporal order judgments at least 75 percent of the time. For tones presented dichotically, Day and Cutting (1971) found that a delay of at least 50 msec is needed to attain 75 percent correct temporal order judgments. The temporal order task seems an intrinsically difficult one but this fact does not in itself explain why not even 90 msec between stimulus onsets was a sufficient interval to give 75 percent correct judgments of stop consonants in the present experiment. This poor performance on temporal order perception for stops in comparison to nonspeech sounds may reflect a difference in functions of the speech and nonspeech processors. In order to report which stop came first or second, the subject must perform a phonetic analysis to identify the stimuli as well as an auditory analysis of temporal order. If the speech processor were incapable

Comparison of the Frequency Distributions of Ear Effect Scores Obtained with Ear Monitoring and Clarity Judgments



Note: A negative score indicates a left-ear advantage and a positive score indicates a right-ear advantage.

Fig. 20

of making such an auditory judgment and if the nonspeech processor were incapable of the phonetic analysis, then information from both speech and nonspeech processors would be required to make temporal order judgments for speech, whereas such coordination between these processors would not be necessary for making temporal order judgments for nonspeech stimuli.

Reporting by ear was easier than reporting by temporal order, but even on ear monitoring the subjects made frequent errors in selection. This result agrees with the findings of Kirstein and Shankweiler (1969) and Halwes (1969) of frequent confusions between ears in the selection of stop consonants presented dichotically with simultaneous onsets at the two ears. Halwes (1969) attributed the difficulty in assigning dichotically presented stops to the correct ear to the absence of distinctive localization cues for each stimulus. Although each ear technically receives a different syllable, in fact, in these experiments the syllables at the two ears are always identical in the vowel portions and thus differ acoustically between ears only in approximately the first 70 msec plus the interval corresponding to the delay between ears. Halwes (1969) argued that because the CV syllables at the two ears are always the same except in the consonant, the apparent localization of both syllables is at the midline. He claimed that subjects experience a fused image of the two syllables, rather than a separate image of each syllable, and thus make frequent errors when asked to report the syllables from one ear.

Whatever the reason for the confusion between ears, the fact that such confusions are frequent makes it clear that instructions to attend to one ear do not prevent the perceptual analysis of material at the unattended ear. This conclusion seemingly contradicts certain results of experiments where dichotic presentation was used as a means of studying selective attention. In those studies, continuous messages were played to opposite ears and subjects were asked to repeat back or "shadow" the message at one ear while ignoring the other. Cherry (1953) reported that subjects easily shadow one of two dichotic messages, and he found that the subjects apparently do not retain any of the content of the unattended message. In order to account for findings of this sort, Broadbent (1958) proposed that the unattended message can be filtered out prior to linguistic analysis on the basis of some physical dimension like ear of arrival. Subsequent research has shown that material on the unattended ear will often be heard correctly if it is prefaced by the subject's name (Moray, 1959) or if it is semantically highly probable within the context of the message at the attended ear (Treisman, 1960). The present study provides further evidence that the ability to ignore one ear depends on the content of the material at the two ears and that a difference in ear of arrival *per se* does not allow the subject to exclude material from awareness. Where the dichotic stimuli are pairs of syllables, the absence of contextual constraints among items at the same ear undoubtedly makes the selection task more difficult than the shadowing of continuous messages.

In addition to evidence from the frequency of confusions in ear monitoring, further evidence that one ear is not being "filtered out" on the ear montoring task comes from a comparison of the results on that task with the results of the temporal order task. In order to report by ear, the listener in theory needs to listen for only one of the two stimuli on each trial, the one at a particular ear. The task of selecting stimuli on the basis of order of arrival requires that the listener detect both stimuli on each trial, even if he need not identify both. A striking finding in these experiments was the similarity of the performance on the two tasks. Although the ear-monitoring

and temporal order-monitoring tasks could be seen as rather dissimilar tasks, scores of individual subjects on three measures of performance--accuracy of monitoring, lag effect, and ear effect--were highly correlated across the two tasks. The finding that the same subjects were good at reporting by ear and by temporal order is very strong evidence that the selection of the response is taking place at the same level in both tasks, namely, after, rather than before, the analysis of the stimuli from both ears. The results also suggest that the ability to separate the dichotically presented stimuli in a selective listening task may be an interesting dimension of individual variation on dichotic listening tasks.

Finally, it was observed that the group functions relating the extent of right-ear advantage and lag advantage to relative onset time were identical for the two tasks. This result indicates, again, that the selection processes are undoubtedly operating at the same level in the two tasks. It would seem that at some stage in the processing the lagging stimulus gains in salience relative to the leading stimulus and the stimulus from the right ear gains in salience relative to the stimulus from the left ear. The selection instructions seem to apply after these alterations in the relative clarity of the two stimuli have occurred.

A comparison of the clarity judgment and ear-monitoring method revealed that the size of the lag effect varied depending on the method but that the size of the right-ear effect was not influenced by the method. The insensitivity of the ear effect to the change in method may be related to the fact that the ear effect operates maximally at the short delay intervals. It was observed in Figure 2 that the longer the delay between syllables, the more often subjects were able to report both syllables correctly. Figure 18 shows that the ear asymmetry arises primarily at short delays where subjects are often unable to identify both stops even when requested to do so. If only one of the stops can be identified, it makes little difference whether the listener is under instructions to report one stop or both. In contrast to the ear effect, the lag effect operates maximally at longer delay intervals (50 to 70 msec) where the subjects in Experiment 1 were found to be more accurate in reporting both syllables.

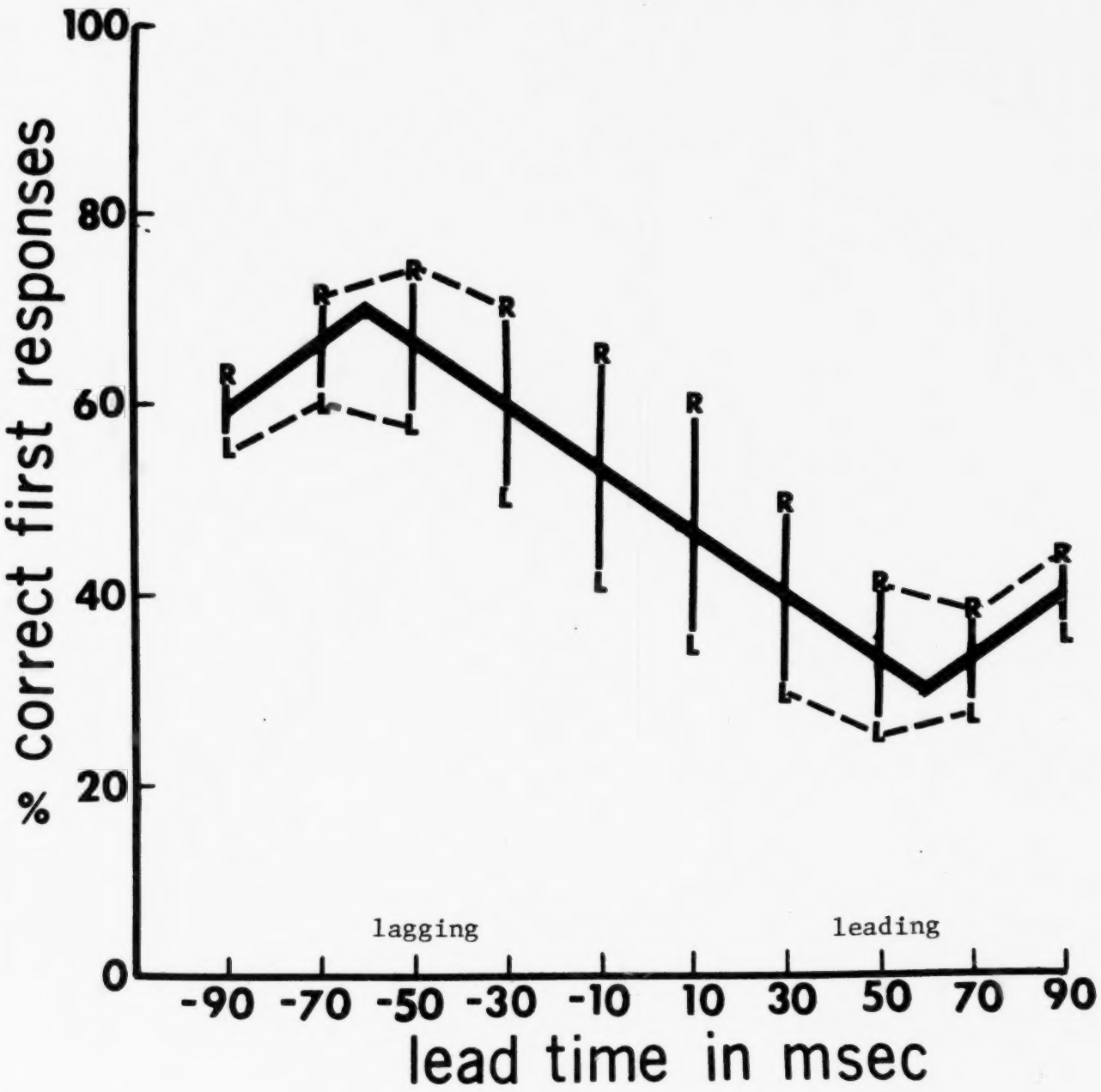
In Experiment 1 there was found to be a difference between ears in the delay which gave the maximum lag effect. The lag effect was greatest for the left ear with a delay of 70 msec, but for the right ear the lag effect was greatest with a delay of 50 msec. No attempt was made in Experiment 1 to explain this difference between ears; however, certain of the results in Experiment 2 suggest a way of accounting for the difference in the location of the peak lag effect for the two ears.

Figure 18 in Experiment 2 shows that the right-ear advantage decreases monotonically with increasing delays between ears. This means that the difference between ears is greater with a 50-msec delay than with a 70-msec delay. If the real peak of the lag effect were at about 60 msec delay rather than at either 50 or 70 msec, the difference in the size of the right-ear effect between the 50- and 70-msec points would create an apparent left-ear peak at 70 msec and an apparent right-ear peak at 50 msec delay.

This reasoning is illustrated in Figure 21 which shows a hypothetical plot of the lag effect, symmetrical around 60 msec, and the right- and left-ear values which could be observed at the 50- and 70-msec delay points. Because the ear effect is greater at 50 than at 70 msec, the left-ear performance declines. Thus, the apparent difference between the location of the peaks for the two ears can be attributed to the consequences of a simple additive relationship between the lag effect and right-ear effect and to the coarseness of the sampling of delay times around the peak interval.

The location of the peak of the lag effect may be significant for understanding why the effect occurs. One possibly important fact about the location of the peak is that 60 msec corresponds very closely to the duration of the stop consonant transitions (the changes in formant frequency in the initial portion of the syllable which provide the acoustic information about the place of articulation of the stops). It may be that the lag effect relates to the processing of transitions. This notion could be tested by experiments which look for a shift in the location of the peak depending upon the duration of the formant transitions. On the other hand, the correspondence between the location of the peak and the duration of the formant transitions could be a coincidence. If, as suggested by Studdert-Kennedy et al. (1970), the lag effect reflects an interruption of phonetic recognition processes, then 60 msec may correspond to a critical interval at which the arrival of another stimulus is most likely to disrupt processing.

Hypothetical Lag Effect Function with an Assumed Peak at 60 msec Delay and the Right- and Left-Ear Curves Which Would Result if a Linearly Decreasing Right-Ear Effect were Added Orthogonally to the Lag Effect at Each Delay Interval



Note: This model accounts for the difference in location of the right- and left-ear lag effect peaks observed in Figure 11.

Fig. 21

EXPERIMENT 3. PERCEPTION OF STOP CONSONANTS AND VOWELS IN DICHOTICALLY PRESENTED CV SYLLABLES

From the results of the selective listening experiments, it seems safe to assume that the lag effect is a genuine perceptual phenomenon. However, it is still not known at what level of perceptual processing the effect originates. It has been thought that the lag effect is peculiar to the perception of encoded speech sounds and reflects the operation of special speech decoding apparatus (Studdert-Kennedy et al., 1970). On the other hand, it is conceivable that the lag effect is a more general phenomenon of auditory perception. One approach toward investigating this issue would be to see whether the lag effect can be obtained for other stimuli besides the stop consonants. Of particular interest are the unencoded speech sounds like steady-state vowels.

There is evidence from dichotic listening studies and from other sources that steady-state vowels can be processed in either the speech mode or the nonspeech mode. When vowels are presented dichotically, the results are quite different from the effects observed with competing stops. For stops, most people give highly reliable right-ear effects. For dichotically presented steady-state vowels, ear effects are typically smaller than ear effects for stops, and more people show left-ear effects for vowels than for stops. Some studies have reported statistically significant right-ear effects for vowels (Chaney and Webster, 1966; Kirstein and Shankweiler, 1969; Weiss and House, 1970), but in other studies, the effects were not significant (Shankweiler and Studdert-Kennedy, 1967; Darwin, 1969; Studdert-Kennedy and Shankweiler, 1970).

There is considerable support for the view that the inconsistent lateralization of steady-state vowels reflects the ability of vowels to be processed either in the speech mode, in the left hemisphere, or in the nonspeech mode, in the right hemisphere. In a recent paper Spellacy and Blumstein (in press) reported that dichotic syllables differing in the vowel gave a left-ear effect when embedded in a test with nonspeech filler items but that the same stimuli gave a right-ear effect when the filler items were words. No context effect was found for dichotically presented stops, which gave a right-ear effect under both conditions. An effect of test context on vowel lateralization was also reported by Darwin (1970). He compared the laterality effect for a set of vowels in a test where all the vowels were from the same vocal tract with the laterality effect obtained for the same vowel pairs embedded in a test where vocal tract size for the other pairs could vary from trial to trial. The same vowel pairs gave no ear effect in a test by themselves but gave a significant right-ear effect in the context of the larger test. Darwin (1970) attributed the shift toward a right-ear effect to the greater complexity of the perceptual task when the vocal tract size varies.

The fact that vowels usually behave in experimental tasks in a manner intermediate between speech and nonspeech sounds made the steady-state vowels seem an appropriate class of sounds to study in connection with the lag effect.

If the lag effect is a phenomenon found only with encoded speech sounds, then the lag effect should not occur or should be much reduced if the stimuli are steady-state vowels. A pilot study by Porter, Shankweiler, and Liberman (1969) gave results consistent with this prediction. They used as stimuli six isolated steady-state vowels, each 400 msec in duration, and presented these vowels for identification in a dichotic test with difference of 0 to 120 msec in the onsets of vowels at the two ears. They tested four subjects, all of whom had lag effects for stops, and found that three of the four were more accurate in identifying the leading vowels than the lagging vowels. One subject had a lag effect for vowels, but this was considerably smaller than the effect seen with stops. The authors interpreted the absence of a lag effect for vowels as evidence that the lag effect is associated with perception in the speech mode.

Up to this point we have learned that dichotically presented syllables contrasting in the initial stop consonant give lag effects, while dichotically presented isolated steady-state vowels apparently give lead effects. The main purpose of Experiment 3 was to determine whether this difference between encoded and unencoded speech sounds in the effects of interaural delay could be observed when stops and vowels were made to contrast between ears within the same syllables. The subjects in Experiment 3 listened to pairs of syllables differing between ears in both the stop and vowel, and they were asked to report either the stops or the vowels. This experiment would show whether the presence of the shared vowel is an essential condition for obtaining a lag effect for stop consonants. The experiment was also expected to show whether the same effects can be obtained for vowels in isolation and in syllabic context.

Method

Test tapes. Experiment 3 required the use of two different tapes. One was the tape described in Experiment 1 where syllables differed in the stop but not in the vowel. The second tape was constructed specifically for Experiment 3. For this tape the same nine syllables described in Experiment 1--[bɛ], [dɛ], [gɛ], [ba], [da], [ga], [bɔ], [dɔ], [gɔ]--were recorded in pairs onto a two-channel tape in such a way that both the consonant and vowel differed between ears on each trial. There are thirty-six possible pairs of this sort; these are listed in Table II. Three delay intervals were used--0, 50, and 70 msec. Each of the thirty-six pairs of syllables appeared twice on the tape at each delay, once with channel 1 delayed and once with channel 2 delayed. The pairs and delays were arranged randomly within the 180-trial tape.

Test conditions. Each of the subjects in this experiment took three different tests.

(1) Condition "C." This condition is an exact replication of the dichotic presentation condition in Experiment 1. Subjects listened to the tape in which stops contrasted between ears and vowels were shared. The instructions to the subjects were to report both stops, guessing if necessary, and to record the clearer stop in the first column of the answer sheet. Responses were limited to "B," "D," and "G."

Table II. Syllable pairs included on the stimulus tape in Experiment 3.

Syllable on Channel 1	Syllable on Channel 2
BE . . .	DA
DA . . .	BE
BA . . .	DE
DE . . .	BA
BE . . .	DO
DO . . .	BE
BO . . .	DE
DE . . .	BO
BA . . .	DO
DO . . .	BA
BO . . .	DA
DA . . .	BO
BE . . .	GA
GA . . .	BE
BA . . .	GE
GE . . .	BA
BE . . .	GO
GO . . .	BE
BO . . .	GE
GE . . .	BO
RA . . .	GO
GO . . .	BA
BO . . .	GA
GA . . .	BO
DE . . .	GA
GA . . .	DE
DA . . .	GE
GE . . .	DA
DE . . .	GO
GO . . .	DE
DO . . .	GE
GE . . .	DO
DA . . .	GO
GO . . .	DA
DO . . .	GA
GA . . .	DO

Each of these combinations occurs with simultaneous onset on the two channels, with channel 1 leading by 50 and by 70 msec and with channel 1 lagging by 50 and by 70 msec.

The letters E, A, and O represent the phonetic qualities [ɛ], [a], and [ɔ], respectively.

(2) Condition "CV-C." In this condition the subjects listened to the tape in which both stops and vowels differed between ears on any trial. The subjects were instructed to report both of the stops on each trial, recording the clearer of the stops in the first column of the answer sheets. The subjects were asked to ignore the vowels and concentrate on the consonants.

(3) Condition "CV-V." The subjects listened again to the tape in which both stops and vowels differed between ears on each trial. In this condition the subjects were instructed to attend to the vowels and to try to ignore the stops. They were to identify both vowels on each trial and to record the vowels in order of clarity on the answer sheet. The letters "E," "A,"and "O" were used to indicate the vowel sounds in the words "bet," "hot," and "law" respectively.

Each condition required a one-hour testing session during which 360 trials were given. The three sessions were held on different days. The subjects were randomly assigned to one of two testing orders: (1) C, CV-C, CV-V, or (2) CV-V, CV-C, C.

Subjects. Twelve subjects took all three test conditions. The subjects were volunteers from the introductory psychology classes at the University of Connecticut and received credit toward a course requirement by participating in the experiment. All were right handed, had normal hearing, and were native speakers of English.

Results

Accuracy of identification. Figure 22 shows the percent correct first and second responses for the three test conditions. There were no significant differences among the three conditions in first response accuracy, but significant differences among the conditions were found in accuracy of second responses (Friedman two-way analysis of variance, χ^2=18.5, 2 d.f $p < .001$). In second responses stop consonants were more often correctly identified when vowels differed between ears than when vowels were shared ($p < .006$, two-tailed sign test). Within the same pair of syllables differing between ears in both the stop and vowel, vowels were more accurately identified than stops ($p < .038$, two-tailed sign test). Very few errors were made in vowel identification in either first or second responses.

Effect of relative onset time on overall percent correct. The experimental procedure provides two possible indicators of the lag effect: (1) more accurate identification of lagging than leading syllables and (2) judgments that lagging syllables are clearer than leading syllables. Figure 23 shows the percent correct report of lagging and leading stimuli considering both first and second responses. The lag effect is clearly present only in condition "C," where vowels are shared at the two ears. The lag effect is barely visible in the overall percent correct for conditions "CV-C" and "CV-V," where vowels contrast between ears.

Effect of relative onset time on clarity judgments. Figure 24 shows the pattern of clarity judgments for the three test conditions. In all conditions, lagging stimuli were judged as clearer than leading stimuli. In order to compare the magnitude of the lag effect in the three conditions, lag effect scores were computed for each subject for each test using the formula (Leading - Lagging)/(Leading + Lagging), where "Leading" and "Lagging" refer to the number of first responses corresponding to leading and lagging stimuli. For condition "C," where only stops varied, the lag effect scores were computed on the

Mean Percent Correct First and Second Responses as a Function of the Interval Between Stimulus Onsets for Conditions C, CV-C, and CV-V

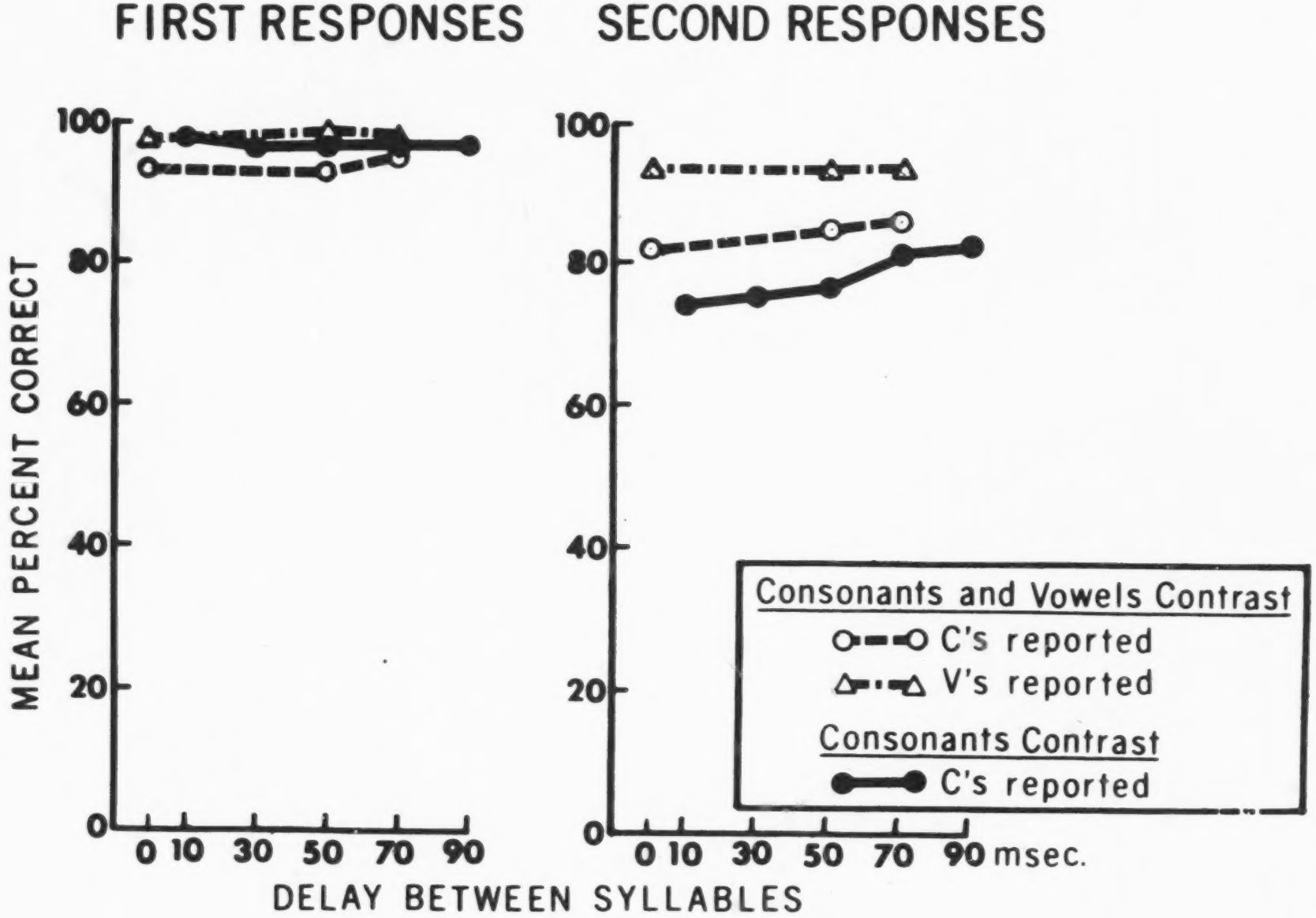


Fig. 22

Mean Percent Correct Responses for Lagging and Leading Stimuli
for Conditions C, CV-C, and CV-V

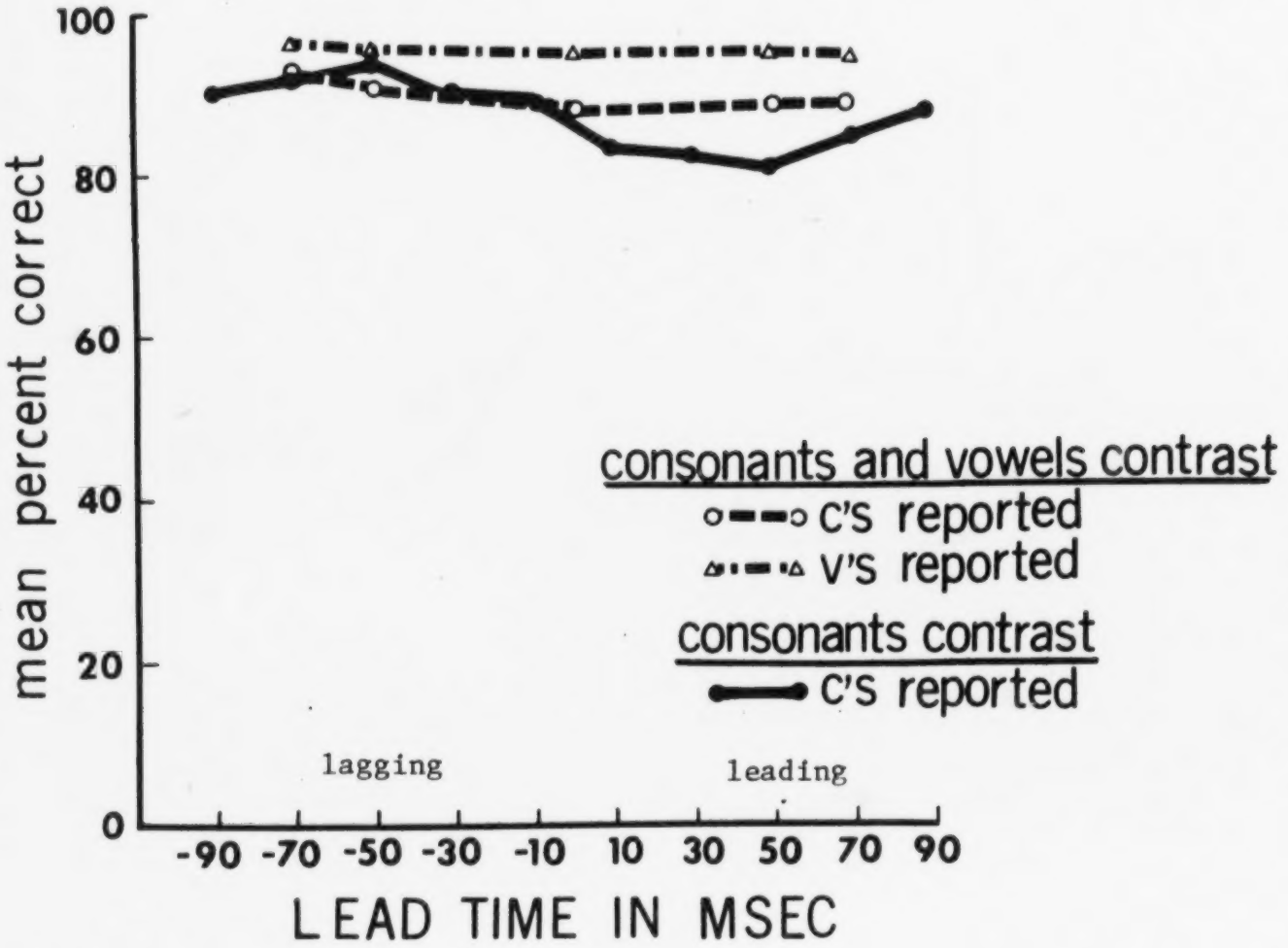


Fig. 23

Mean Percent Correct First Responses Corresponding to Lagging and Leading Stimuli for Conditions C, CV-C, and CV-V

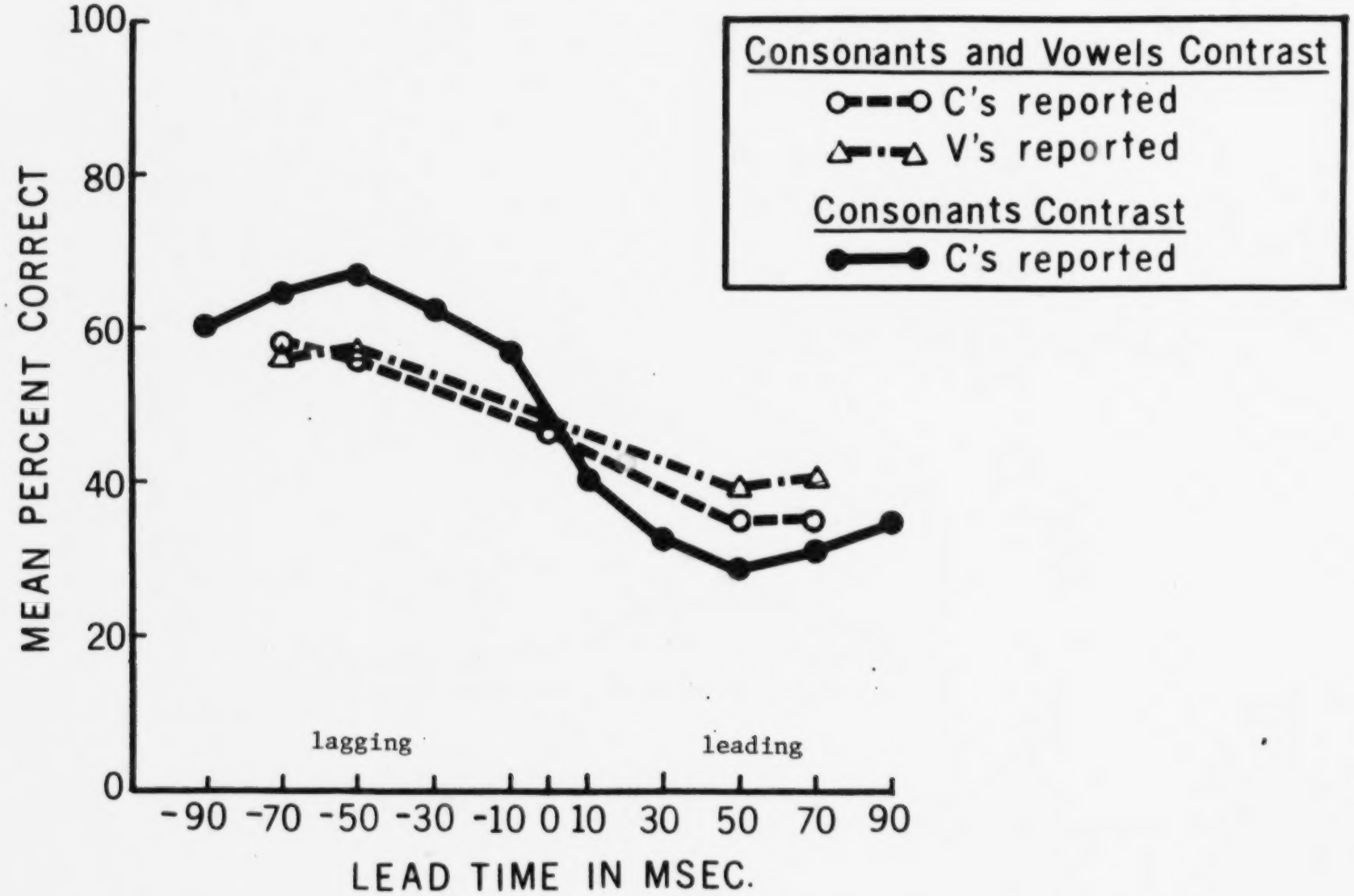


Fig. 24

basis of the 50- and 70-msec delay trials, since these were the delay intervals which the three conditions had in common. A Friedman two-way analysis of variance (Siegel, 1956) indicated that these scores differed significantly among the three conditions (χ^2=9.5, $p < .01$, 2 d.f.). The lag effect in clarity judgments for stop consonants was significantly greater when vowels were shared than when vowels contrasted between ears (C>CV-C, $p < .006$ by a two-tailed sign test). However, there was no significant difference in the lag effect for the CV-C and CV-V conditions.

Of the twelve subjects, eleven had a lag effect in clarity judgments in condition "C" and all twelve had lag effects in conditions "CV-C " and "CV-V." Individual differences in lag effect scores were highly correlated across the three conditions. The Spearman rank correlation coefficients indicating the consistency in individual differences across conditions are shown in Table III.

Table III. Individual differences in the magnitude of the lag effect correlated across the C, CV-C, and CV-V conditions (Spearman rank correlation coefficients).

Conditions	r_S	
C and CV-C	.951	p<.001
C and CV-V	.811	p<.01
CV-C and CV-V	.713	p<.01

Laterality effects. Right-ear effects in clarity judgments were shown to ten subjects in condition "C" and eight in condition "CV-C" and "CV-V." There were no significant differences among the three conditions in the size of the ear effect. Individual differences in the size of the right-ear effect were positively correlated across the three conditions (Table IV).

Table IV. Individual differences in the magnitude of right-ear effect correlated across the C, CV-C, and CV-V conditions (Spearman rank correlation coefficients).

Conditions	r_S	
C and CV-C	.783	p<.02
C and CV-V	.537	p<.10
CV-C and CV-V	.678	p<.10

It is interesting that the two conditions involving stop consonant identification (C and CV-C) correlate more highly with each other than either of them correlates with the vowel identification test. This was true for the lag effect correlations as well as the ear effect.

Discussion

In an earlier study, Porter et al. (1969) reported that isolated steady-state vowels did not give a lag effect. One of the purposes of Experiment 3 was to see whether such a dissociation of effects could be observed for consonant and vowel segments within the same syllable. However, when both stops and vowels were made to differ between ears, a lag effect of approximately the same extent was found whether subjects were attending to the stops or to the vowels. The lag effect observed in first responses when both stop and vowel contrasted could be described as intermediate between that obtained for stops when vowels were shared and the effect obtained for isolated vowels.

In trying to account for this result, one of the factors which seems relevant is the effect of the vowel contrast on overall accuracy of identification. Significantly more stop consonants were correctly identified when vowels differed between ears than when vowels were shared. This rise in performance may be a consequence of a reduction in fusion of stimuli from the two ears when vowels contrast. With a vowel contrast, the dichotic syllables differ acoustically for a full 350 msec, not just for the duration of the stop consonant transitions. It may, therefore, be easier to localize the syllables by ear when vowels contrast than when vowels are shared. One might speculate that this superior spatial separation of stimuli facilities the retention of an auditory image of one syllable while the other is being decoded and thus causes an increase in the number of stimuli correctly identified.

It seems reasonable to suppose that the factors underlying the improved performance also underlie the reduction in the lag effect for stops when vowels contrast. Although in principle the extent of lag effect in clarity judgments is not constrained by the performance level, it seems likely that clarity judgments become more difficult and hence less reliable when both of the stops can be correctly identified.

In the case of the vowel identification task, the main result to be explained is why there was a lag effect rather than a lead effect. There are several reasons why vowels in syllables and vowels in isolation might give different results. When vowels are embedded in CV syllables where the stop consonants differ between ears, the listeners might find it difficult to separate their judgments about the relative clarity of the two vowels from their perception of the stops. Another possibility is that the presence of stop consonants may induce a "set" to process stimuli in the speech mode. The lag effect for vowels could reflect the fact that they are being processed in the speech perception mode, whereas the lead effect for isolated vowels might be associated with perception in the nonspeech mode. If lag effects can be obtained for vowels when they are being perceived in the speech mode, then it should be possible to obtain a lag effect for isolated vowels as well as vowels in syllables under conditions which induce a "set" toward the speech mode. Certain findings in Experiment 4 uphold this prediction.

EXPERIMENT 4. EFFECTS OF DELAY BETWEEN EARS ON THE PERCEPTION OF DICHOTICALLY PRESENTED ISOLATED STEADY-STATE VOWELS

In Experiment 3 it was found that when listeners were requested to judge the relative clarity of vowels in dichotically presented CV syllables where the consonant as well as the vowel differed between ears, all listeners judged the lagging vowels as clearer than the leading vowels. This finding of a lag effect for dichotically presented vowels was unexpected in light of the report of Porter et al. (1969) of a lead effect for vowels. The discrepancy between the findings reported in Experiment 3 and those of Porter et al. (1969) might be related to the fact that the latter study used isolated vowels in syllable context. However, there was some doubt about the reliability of the lead effect for isolated vowels. Porter et al. (1969) described only four subjects, and one of those four actually had a lag effect rather than a lead effect for isolated vowels. Experiment 4 examined the perception of dichotically presented isolated vowels with more subjects in order to determine whether the effects of interaural delay on vowel perception are truly different for vowels in isolation and in syllabic context.

Method

Stimuli. The stimulus set consisted of three vowels whose phonetic qualities were [ɛ], [a] and [ɔ]. Each vowel was 350 msec in duration. The three vowels were acoustically identical to the steady-state vowel portions of the CV syllables used in Experiment 3.

Test construction. The procedures of test construction and counterbalancing were identical to those described in Experiment 1. A 180-item dichotic tape was prepared consisting of pairs of vowels with delays of 10, 30, 50, 70, or 90 msec between vowel onsets at the two ears. Two runs through the tape, with the headphones reversed on the second run, gave a total of 360 trials for each subject, 72 trials at each delay.

Instructions. The instructions to the subjects were similar to those given in Experiments 1 and 3. The subjects were told that they would receive two different vowels on each trial, one vowel at each ear. They were to report both vowels on each trial, guessing if necessary, and to record the clearer of the vowels in the first column of the answer sheet. The letters "E," "A," and "O" were used by the subjects to designate the vowel sounds in the words "bet," "hot," and "law."

Subjects. Twenty right-handed subjects took the isolated vowel test. In the data analysis the subjects were divided into two groups on the basis of their previous experience in dichotic tests. Ten of the subjects had never previously taken part in a dichotic listening experiment and are referred to as "naive" subjects. The other ten subjects had previously taken the CV-V test described in Experiment 3, and seven of them had taken all three of the conditions in Experiment 3; these ten are referred to as "experienced" subjects.

Results

Comparison of clarity judgments of naive and experienced subjects. Because errors in vowel identification were rare, the effects of order or ear of arrival could not be seen in the overall accuracy of identification, but such effects could be observed in clarity judgments. Figure 25 compares first responses for naive and experienced subjects. One is struck by the obvious difference between the two groups. The naive subjects tended to perceive the leading vowel as clearer than the lagging vowel, whereas the experienced subjects had a consistent lag effect. Moreover, the naive subjects showed only a slight right-ear effect whereas the experienced subjects had a much larger right-ear effect. Thus, the naive subjects performed in a manner similar to the subjects described by Porter et al. (1969) for isolated vowels, but the behavior of the experienced subjects resembled that observed in Experiment 3 for vowels in CV syllables.

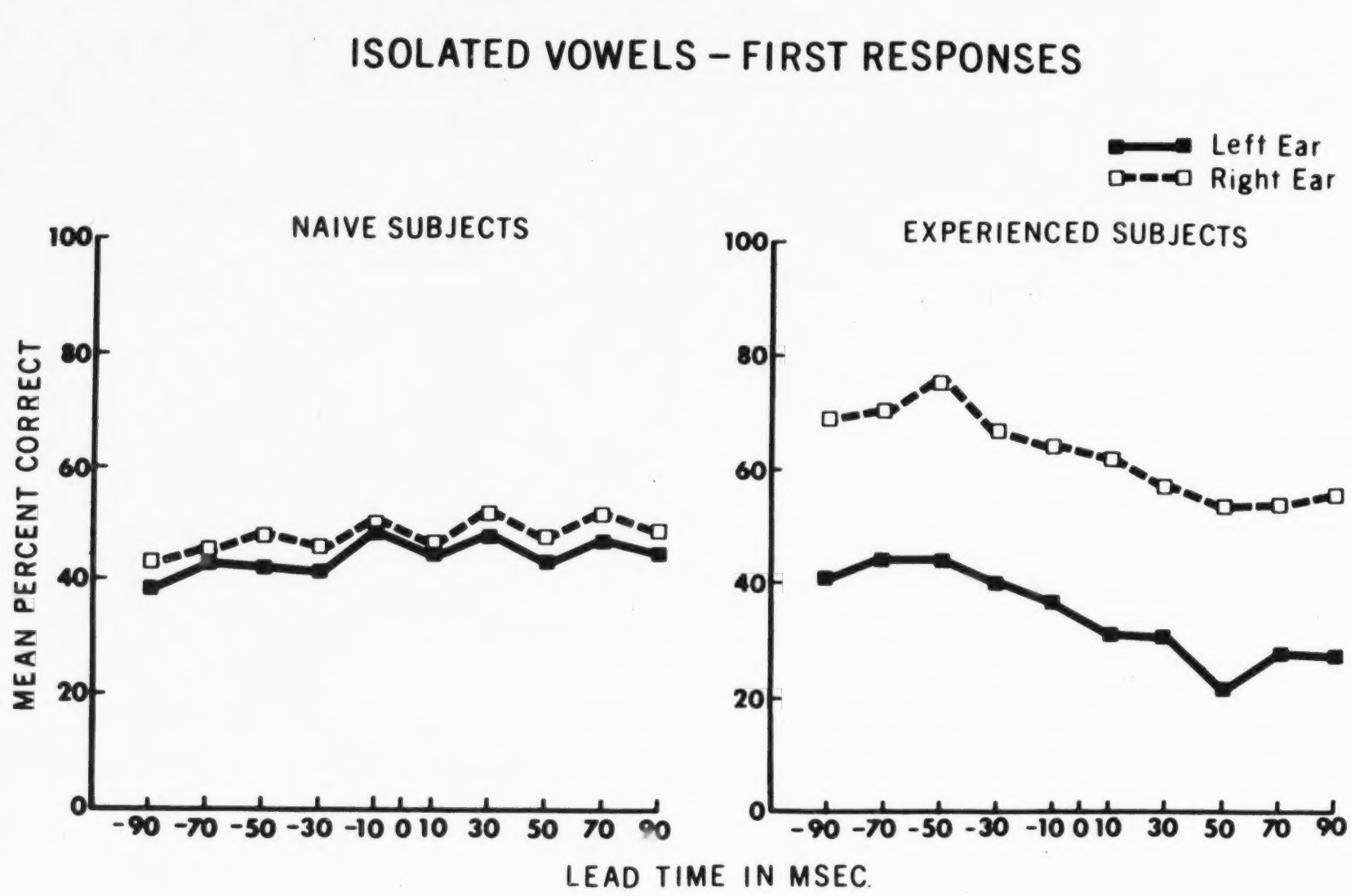
Relationship between the lag effect and right-ear effect. Figure 25 shows an association between the lag effect and right-ear effect; the experienced subjects had a large lag effect and a large right-ear effect, but the naive subjects had a lead effect and a much smaller right-ear effect. The question arose as to whether the association between the two effects could be seen only in a comparison of the two groups or whether this association reflects a more general correlation between the lag effect and right-ear effect which might be observed even within each group.

In order to measure the correlation between the lag effect and the right-ear effect within each group, an index of the magnitude of the lag effect and ear effect was computed for each subject as in the previous experiments. The index of the right-ear effect was (Right - Left)/(Right + Left) where "Right" and "Left" refer to the number of first responses corresponding to stimuli presented to the right or left ears. The index of the lag effect was (Lagging - Leading)/(Lagging + Leading) where "Lagging" and "Leading" refer to the number of first responses corresponding to the lagging or leading stimuli. The Spearman rank correlation coefficient (r_S) was used to measure the extent of association between the lag effect and right-ear effect within each group. For the group of naive subjects there was a significant positive correlation between the lag effect and right-ear effect indices ($r_S = +.71$, $p < .025$). For the group of experienced subjects the correlation between the two effects was positive but of borderline significance ($r_S = +.50$, $p < .10$).

The positive correlation between the right-ear effect and lag effect can be seen in Figure 26 which plots each subject's ear effect on the x-axis against his lag effect on the y-axis. This scatter plot also displays clearly the differences between the naive and experienced group. The naive subjects tend to cluster around the (0,0) point whereas the experienced subjects are more variable and show right-ear effects as large as those typically observed for stop consonants.

Relationship between the lag effect and right-ear effect for the C, CV-C, CV-V, and isolated vowel tests. In Experiment 1 it was reported that for stop

Mean Percent Correct First Responses by Ear for "Naive" and "Experienced" Subjects on the Isolated Vowel Test



Note: Identical onset time conditions are compared for the two ears. Negative lead times refer to lagging stimuli and positive times refer to leading stimuli.

Fig. 25

Scatter Plot Showing the Relation Between an Individual's Lag Effect Score (Y-Axis) and Ear Effect Score (X-Axis) for the Isolated Vowel Test

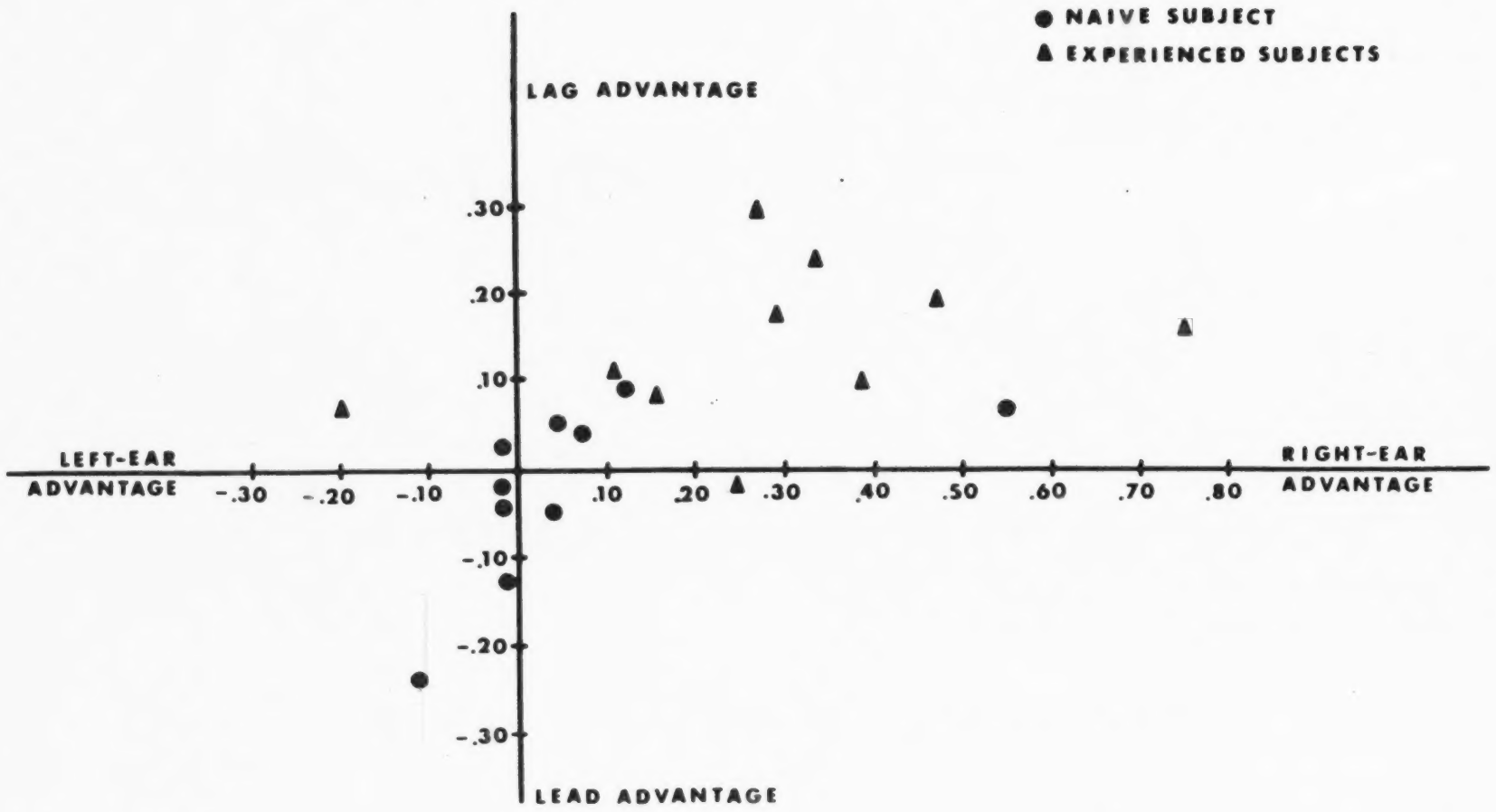


Fig. 26

consonants the magnitude of the lag effect was not significantly correlated with the magnitude of the right-ear effect, and the sign of the correlation between the two effects was negative. However, for isolated vowels there is clearly a significant positive relationship between the lag effect and right-ear effect. In view of these contradictory results for stops and vowels, a more general survey was made of the incidence of lag effects and right-ear or left-ear effects and of the relationship between the ear effect and lag effect for the various test conditions described in Experiments 1, 3, and 4. Table V summarizes these data for each of the four test conditions: C, CV-C, CV-V, and isolated vowels. Only naive subjects are considered in this tabulation, that is, subjects who had not taken any dichotic test prior to the test in question.

Table V shows that for the condition in which only stop consonants differed between ears nearly all of the naive subjects had both right-ear effects and lag effects. Subjects with left-ear effects also had lag effects in this condition, and the correlation between the ear effect and lag effect was negative and not significant.

For the isolated vowel condition, the incidence of right-ear effects and of lag effects was reduced in comparison with stop contrast condition. Also, for isolated vowels, the subjects who had left-ear effects tended not to have lag effects, and there was a significant positive correlation between the extent of right-ear advantage and of lag effect.

For the conditions in which both stops and vowels differed between ears, the pattern of results depended upon whether the subjects were attending to stops (CV-C) or vowels (CV-V). Naive subjects identifying vowels in syllables resembled the naive subjects on the isolated vowel tests; both conditions showed a lower incidence of right-ear effects than the two conditions in which stops were being identified, and the extent of right-ear effect for vowels in syllables was also positively correlated with the extent of the lag effect. When stops were being identified (CV-C), the correlation between the magnitude of the ear effect and lag effect was negative and nonsignificant, as was the case when vowels were shared and only stops differed between ears.

Comparison of lag effects for the C, CV-C, CV-V, and isolated vowel tests. There were seven subjects who took the three conditions in Experiment 3 as well as the isolated vowel test. Because the isolated vowel test was taken after the other three tests, these subjects were included among the "experienced" subjects in the isolated vowel condition. Figure 27 compares the lag effect for these seven subjects in the four conditions. Percent correct first responses at each delay interval are averaged over the two ears. All three conditions in which vowels contrasted between ears (CV-C, CV-V, isolated vowels) had lag effects of approximately the same extent, regardless of whether stops or vowels were being identified. These three conditions gave a smaller lag effect than the condition in which vowels were shared at the two ears. The location of the peak in the lag effect is not marked as clearly with the isolated vowels as with the stops in condition C, but the peak appears to be in about the same place for the two conditions, that is, between 50 and 70 msec delay. More data are necessary to establish the location of the peak for vowels with greater confidence.

TABLE V. Comparison of the C, CV-C, CV-V, and isolated vowel conditions for naive subjects. The incidence of lag effects, the incidence of right-ear effects, and the relation between the ear effect and lag effect.

Condition	N*	Right-ear effects	Lag effects	Right-eared subjects with lag effects	Left-eared subjects with lag effects	Correlation (r_S) between the lag and ear effects	
C	17	89%	94%	94%	100%	-.37	n.s.
CV-C	10	80%	90%	87%	100%	-.31	n.s.
CV-V	20	40%	80%	100%	67%	+.67	$p < .01$
Isolated vowels	10	50%	50%	80%	20%	+.71	$p < .02$

*Some of the subjects in the CV-V condition and all the subjects in the CV-C condition were run specifically for this tabulation and were not included among the subjects previously described in Experiments 3 and 4.

Mean Percent Correct First Responses as a Function of Lag or Lead Time
for Seven Subjects on the C, CV-C, CV-V, and Isolated Vowel Tests

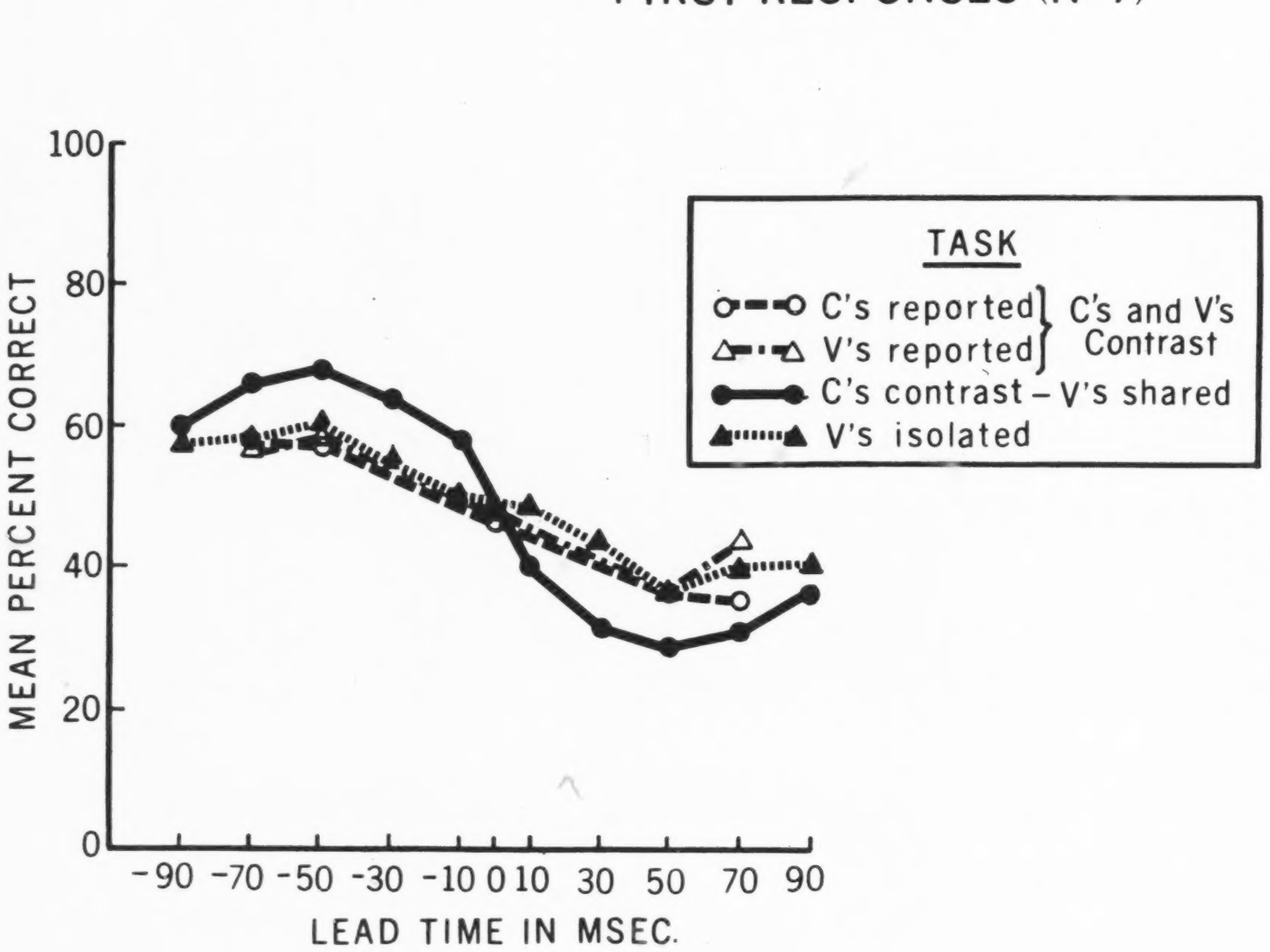


Fig. 27

172

Discussion

Experiment 4 showed that the presence of stop consonants is not an essential condition for obtaining the lag effect. The effect can be obtained for steady-state vowels isolated from syllabic context as well as for vowels and stops in CV syllables. While a lag effect was found for vowels as well as stops, there were, nevertheless, differences between the effects obtained for stops and vowels in the dichotic listening tasks. For vowels, in contrast to stops, there was an association between the lag effect and the right-ear effect. This association manifested itself in two ways. First, previous experience with other dichotic tests seemed to produce a shift toward both the right-ear advantage and lag advantage for vowels. Second, regardless of previous experience, the extent of lag effect and the extent of right-ear effect were positively correlated within each group of subjects on the vowel identification tasks. These findings with vowels were very different from the findings with stop consonants. For stops, no previous experience in other dichotic tests was needed in order to obtain both the lag effect and the right-ear effect. Moreover, for stops there was no significant correlation between the two effects within a group of subjects. For stops, the few subjects who had left-ear effects also had lag effects, but the frequency with which lag effects accompanied left-ear effects was much reduced for vowels.

The results of Experiment 4 agree with those of earlier studies in showing that perceptual effects obtained for dichotically presented vowels are more labile than effects obtained for stops and are more susceptible to experimental manipulation. The experiments of Spellacy and Blumstein (in press) and Darwin (1970) showed that vowels will give either a right-ear effect like stops or a left-ear effect like nonspeech sounds depending on the test context in which critical vowel pairs were embedded. The present experiment indicates that a subject's previous experience in other tests may also affect the direction of lateralization of dichotic vowels: 90 percent of the subjects with previous experience in dichotic speech tests but only 50 percent of the naive subjects had right-ear effects for isolated vowels.[5] These shifts in lateralization of dichotically presented vowels may be plausibly interpreted as reflecting alternate processing of vowels by speech and nonspeech perceptual mechanisms.

The question of the relationship between the lag effect and the right-ear effect is of interest for two reasons. First, because the lag effect and right-ear effect occur within the same experimental task and because both effects involve a suppression of one stimulus relative to the other, it is conceivable that the lag effect and laterality effect are really different facets of a single phenomenon. However, several of the findings in these experiments

---

5 There may have been an effect of previous experience on vowels in CV syllables as well as on isolated vowels, but there were not enough experienced subjects in the CV-V condition to resolve the question.

point to the conclusion that the lag effect and ear effect should be considered separate phenomena. One relevant finding from Experiment 2 is that the time course of the two effects is quite different; the ear effect is greatest with simultaneous or near simultaneous onsets and decreases with longer delays between stimuli, but the lag effect rises between 10 and 60 msec delay. When a comparison was made of the effects of relative onset time for the left and right ears, it was found that the ear effect seemed to add orthogonally to the lag effect. This result would appear to indicate that we are dealing with two different phenomena.

Second, the relationship between the lag effect and ear effect is of interest because it bears on the question of whether or not the lag effect is associated with perception in the speech mode. Although the lag effect may be essentially independent of the ear effect, the lag effect might, nevertheless, also reflect the operation of speech decoding mechanisms. The presence of a right-ear effect is relevant to evaluating this hypothesis because the right-ear effect is an independent indicator of processing in the speech mode. The finding of a right-ear effect for a particular set of stimuli is evidence that these stimuli are being analyzed by special speech processors in the left hemisphere. The positive correlation between the magnitude of the lag effect and right-ear effect for vowels is strong evidence that the lag effect is associated with speech perception. The absence of any correlation between the lag effect and ear effect for stops does not invalidate this conclusion from the vowel data; the presence of a positive correlation for vowels but not for stops can be explained if it is realized that the direction of lateralization is of different significance for stops and vowels.

Because stop consonants are highly encoded speech sounds, it seems reasonable to assume that stops are always processed by special speech decoding mechanisms. The direction of ear asymmetry for dichotically presented stops may be interpreted as an indicator of the hemispheric location of the cerebral speech areas within that subject. Subjects with left-ear effects for stops presumably have right-hemisphere language "dominance." If the lag effect accompanies perception in the speech mode, one would expect to observe a lag effect for stops regardless of the direction of the ear effect.

The meaning of a left-ear effect for vowels is ambiguous because vowels can be processed in either the speech mode or the nonspeech mode. A left-ear effect for vowels may mean: (1) that the subject has normal, left-hemispheric language representation and is analyzing vowels in the speech mode, or (2) that the subject has right-hemispheric language representation and is analyzing vowels in the speech mode. Since most people have left-hemispheric language representation, the first account of the left-ear effect for vowels will be correct in most instances. If the lag effect occurs only when vowels are being processed in the speech mode, the lag effect would be associated with a left-ear effect for vowels only rarely, in the second of the above situations. Thus, if the lag effect and right-ear effect reflect independent processes within the speech mode, a positive correlation between the two effects would be expected for sounds like vowels, which are processed in either the speech or the nonspeech mode, but no correlation would be expected for highly encoded sounds like stops, for which special speech decoding is obligatory.

SUMMARY

Four experiments were described which were all concerned with the perception of dichotically presented speech sounds. Experiment 1 demonstrated that when stop consonant-vowel syllables differing in the stop were separated in onset time between ears by 10, 30, 50, 70, or 90 msec, the lagging stop consonants were judged as clearer than the leading stops. This "lag effect" was shown to be specific to the dichotic mode of presentation. When the competing syllables were delivered to the same ear (monotic presentation), the leading stops were more accurately identified than the lagging stops. The monotic lead effect is probably attributable to peripheral masking, but the dichotic lag effect is clearly central in origin. It was suggested that the lag effect might reflect the operation of special speech processing mechanisms. This notion seemed reasonable on theoretical grounds because there is good evidence that speech sounds and nonspeech sounds are processed by different mechanisms in different areas of the brain. The main goal of the research was to provide evidence relevant to the evaluation of the hypothesis that the lag effect is a speech perception phenomenon.

The first issue to be decided was whether the lag effect was to be viewed as a genuine perceptual phenomenon. In Experiment 1 the subjects had been asked to report both stops on each trial and to judge which of the stops sounded clearer. The main evidence of the lag effect was the finding that lagging stops were judged as clearer than leading stops. This result might reflect a response bias rather than an actual perceptual advantage for lagging syllables. However, Experiment 2 gave convincing evidence that the lag effect is a perceptual effect. It was found in Experiment 2 that when subjects were instructed to report only the lagging syllable on each trial, they were more accurate than when they were instructed to report only the leading syllable. The subjects also proved to be inaccurate in judging the order of arrival of the syllables, so it would be implausible to relate the lag effect to the conscious perception of temporal order of arrival. In another condition in Experiment 2 it was found that subjects also showed a lag effect when they were asked to report the syllables arriving at a particular ear. The magnitude of the lag effect was exactly the same regardless of whether the instructions were to report syllables by order of arrival or ear of arrival. The results of Experiment 2 demonstrated that the lag effect is a robust perceptual phenomenon which does not depend on any particular recall strategy.

Although Experiment 2 pointed to a perceptual origin of the lag effect, it was still unclear whether the effect was specific to the perception of encoded speech sounds like the stop consonants or whether the effects observed with stops were merely an instance of a more general auditory phenomenon. There is a considerable experimental literature which shows that encoded speech sounds and nonspeech sounds give very different results on certain perceptual tasks. Steady-state vowels have been found to give results intermediate between the results for speech and nonspeech sounds. This has been interpreted to mean that vowels may be analyzed in either (or both) the speech and nonspeech

modes, an interpretation which seems reasonable when one considers the acoustic nature of vowels, sustained resonances rather like musical chords.

If the lag effect is associated with perception of encoded speech sounds, then it would be expected that vowels either would not give a lag effect or would show the effect to a reduced extent. The perception of dichotically presented time-staggered vowels was examined in Experiments 3 and 4. In Experiment 3 subjects listened to dichotically presented CV syllable pairs contrasting between ears in both the stop consonant and vowel. In one condition the subjects were to attend only to the consonants and ignore the vowels; in the other condition they were to attend only to the vowels and ignore the consonants. In both conditions they were to identify both of the competing sounds on each trial and to judge which sounded clearer. It was found that for both the consonant and vowel tasks, lagging stimuli were judged as clearer than leading stimuli. However, in both these cases the magnitude of the lag effect was much less than that obtained when only stops contrasted while vowels were shared. Thus, it was found in Experiment 3 that causing vowels to contrast between ears reduced the size of the lag effect, but there was no difference in the magnitude of the effect depending on whether stops or vowels were being identified.

Experiment 4 investigated the perception of steady-state vowels in isolation from syllabic context. It was found that a subject's performance on this task varied depending on whether or not he had had previous experience in tasks involving the identification of vowels in CV syllables. The subjects who had on some previous occasion taken the test in which they identified dichotic vowels in CV syllables had lag effects on the isolated vowel test. However, subjects who had never previously taken a dichotic test showed no particular preference for either the lagging or leading vowel.

The results on the vowel experiments were seen as consistent with the view that the lag effect is related to speech decoding processes. Results of other experiments have demonstrated that vowels can be shifted into or out of the speech mode depending on the experimental conditions in which vowels were tested. Subjects who were naive to speech perception experiments did not have a lag effect for isolated vowels, but subjects with experience in speech experiments did have a lag effect. The previous experience could be viewed as inducing a "set" to process the vowels in the speech mode. Likewise, embedding the vowels in CV syllables could also be seen as a means of inducing subjects to perceive vowels in the speech mode. This interpretation relating the lag effect for vowels to processing in the speech mode and the absence of a lag effect to processing in the nonspeech mode is made extremely convincing by other analyses relating the dichotic lag effect to the dichotic right-ear effect.

When syllables differing in the stop consonant are presented simultaneously to opposite ears, syllables at the right ear are on the average more accurately reported than those at the left ear. The right-ear advantage in dichotic listening has been attributed to the functional asymmetry of the two cerebral hemispheres of the human brain. The left hemisphere is more important than the right for the processing of linguistic material, including speech sounds. The right cerebral hemisphere appears to be more important than the left for processing nonverbal auditory material. Because the right ear has a stronger

representation at the left hemisphere than the left ear, speech sounds presented to the right ear are more accurately perceived than those presented simultaneously to the left ear. For dichotically presented nonspeech sounds the ear effect is reversed, with the left ear being superior to the right ear.

Because the right-ear effect is obtained only for stimuli which are being processed by the speech processors in the left hemisphere, the right-ear effect is an independent indicator that stimuli are being processed in the speech mode. In the experiment where dichotically presented stop consonants are staggered in time at the two ears, both the right-ear effect and the lag effect can be observed. If the lag effect is a speech perception phenomenon like the right-ear effect, some interaction between the lag effect and the right-ear effect might be expected.

There is ample evidence that the lag effect and right-ear effect are to be viewed as separate phenomena; that is, one of these effects cannot simply be seen as a special case of the other. For example, the right-ear effect and lag effect have different time courses. The right-ear advantage is largest when stimulus onsets at the two ears are simultaneous and decreases monotonically with increasing delay between ears. The lag effect, although visible with a 10-msec delay between stimulus onset, increases in size with still longer delays. The advantage for the lagging syllable is maximal with a 60-msec delay between stimulus onsets at the two ears. With delays greater than 60 msec the lag effect begins to decline. Further evidence of the independence of these effects was the finding that the frequency of report from the left and right ears at any delay interval could be predicted by simply adding the lag effect and ear effect to each other. That is, the effect of relative onset time was the same for the two ears.

Another way of trying to relate the lag effect to the ear effect was to see whether there was any correlation between the size of the lag effect and the size of the ear effect for individual subjects. For the tests involving stop consonant identification, no significant correlation was seen in the magnitude of the ear effect and lag effect. However, for the tests involving the identification of vowels, high positive correlations were found relating the extent of right-ear effect to the extent of lag effect. These correlations were significant in two independent groups of subjects on the vowel tests. The highest correlation was observed within the naive subjects in the isolated vowel condition. A significant positive correlation was also observed for vowels in CV syllables. For the experienced subjects on the isolated vowel test the correlation was positive but not significant.

These findings were interpreted as showing that the lag effect and right-ear effect are independent phenomena within the speech mode. Stop consonants, which are always decoded in the speech mode, would show both an ear effect and a lag effect; but if the two effects are essentially independent within the speech mode, no correlation in the magnitude of the lag effect and right-ear effect would be expected for stops. For vowels, a correlation between the lag effect and ear effect scores would be expected because vowels can be processed alternately in the speech or nonspeech modes. Both the lag effect and right-ear effect would be expected on trials when the stimuli were being processed in the speech mode, but neither of these effects would be expected on trials when the stimuli were being processed in the nonspeech mode. Thus,

the finding of a positive correlation between lag effect and ear effect scores for vowels but not for stops supports the view that the lag effect and right-ear effect are separate phenomena both of which relate to processing in the speech mode.

BIBLIOGRAPHY

Berlin, C.I., Loovis, C.F., Lowe, S.S., Cullen, J.K. and Thompson, C. (1970) Dichotic and monotic time-staggered speech perception. Paper read at the 79th Meeting of the Acoustical Society of America, Atlantic City, April 1970.

Borkowski, J.G., Spreen, O. and Stutz, J.Z. (1965) Ear preference and abstractness in dichotic listening. Psychon. Sci. 3, 547-548.

Broadbent, D.E. (1958) Perception and Communication. (Pergamon Press, New York).

Chaney, R.B. and Webster, J.C. (1966) Information in certain multidimensional sounds. J. Acoust. Soc. Amer. 40, 447-455.

Cherry, E.C. (1953) Some experiments on the recognition of speech, with one and with two ears. J. Acoust. Soc. Amer. 25, 975-979.

Cooper, F.S. and Mattingly, I.G. (1969) Computer-controlled PCM system for investigation of dichotic speech perception. J. Acoust. Soc. Amer. 46, 115(A).

Cooper, F.S., Rand, T.C., Music, R.S. and Mattingly, I.G. (1971) A voice for the laboratory computer. IEEE International Convention Digest, 104-105.

Curry, F.K.W. (1967) A comparison of left-handed and right-handed subjects on verbal and non-verbal dichotic listening tasks. Cortex 3, 343-352.

Curry, F.K.W. and Rutherford, D.R. (1967) Recognition and recall of dichotically presented verbal material by right- and left-handed persons. Neuropsychologia 5, 119-126.

Darwin, C.J. (1969) Auditory perception and cerebral dominance. Unpublished Ph.D. thesis, University of Cambridge.

Darwin, C.J. (1970) Ear differences in recall of vowels produced by different sized vocal tracts. Paper read at the 79th Meeting of the Acoustical Society of America, Atlantic City, April 1970.

Darwin, C.J. (1971a) Dichotic forward and backward masking of speech and nonspeech sounds. Paper read at the 81st Meeting of the Acoustical Society of America, Washington, D.C., April 1971.

Darwin, C.J. (1971b) Ear differences in the recall of fricatives and vowels. Quart. J. Exp. Psychol. 23, 46-62.

Day, R.S. (1968) Fusion in dichotic listening. Unpublished Ph.D. thesis, Stanford University.

Day R.S. (1969) Temporal order judgments in speech. Paper read at the 9th Annual Meeting of the Psychonomic Society, St. Louis, November 1969.

Day, R.S. and Cutting, J.E. (1971) What constitutes perceptual competition in dichotic listening? Paper presented to the Eastern Psychological Association, New York City, April 1971.

Dirks, D.D. (1964) Perception of dichotic and monaural verbal material and cerebral dominance for speech. Acta Otolaryngologica 58, 73-80.

Eimas, P.D. (1963) The relation between identification and discrimination along speech and nonspeech continua. Language and Speech 6, 206-217.

Haggard, M.P. (1969) Perception of semi-vowels and laterals. J. Acoust. Soc. Amer. 46, 115(A).

Halwes, T.G. (1969) Effects of dichotic fusion in the perception of speech. Unpublished Ph.D thesis, University of Minnesota.

Hirsch, I.J. (1959) Auditory perception of temporal order. J. Acoust. Soc. Amer. 31, 759-767.

Inglis, J. (1965) Dichotic listening and cerebral dominance. Acta Otolaryngologica 60, 231-238.

Kimura, D. (1961a) Some effects of temporal-lobe damage on auditory perception. Canad. J. Psychol. 15. 156-165.

Kimura, D. (1961b) Cerebral dominance and the perception of verbal stimuli. Canad. J. Psychol. 15, 166-171.

Kimura, D. (1964) Left-right differences in the perception of melodies. Quart. J. Exp. Psychol. 16, 355-358.

Kimura, D. (1967) Functional asymmetry of the brain in dichotic listening. Cortex 3, 163-178.

Kirstein, E. and Shankweiler, D. (1969) Selective listening for dichotically presented consonants and vowels. Paper read before the 40th Annual Meeting of the Eastern Psychological Association, Philadelphia, April 1969.

Liberman, A.M. (1971) The grammars of speech and language. Cognitive Psychology 1, 301-323.

Liberman, A.M., Cooper, F.S., Shankweiler, D.P. and Studdert-Kennedy, M. (1967) Perception of the speech code. Psychol. Rev. 74, 431-461.

Liberman, A.M., Cooper, F.S., Studdert-Kennedy, M., Harris, K.S. and Shankweiler, D.P. (1966) Some observations on the efficiency of speech sounds. Paper presented at the XVIII International Congress of Psychology, Moscow, August 1966.

Liberman, A.M., Harris, K.S., Kinney, J.A. and Lane, H. (1961) The discrimination of relative onset-time of the components of certain speech and nonspeech patterns. J. Exp. Psychol. 61, 379-388.

Lowe, S.S., Cullen, J.K., Thompson, C., Berlin, C.I., Kirkpatrick, L.L. and Ryan, J.T. (1970) Dichotic and monotic simultaneous and time-staggered speech. J. Acoust. Soc. Amer. 47, 76(A).

Massaro, D.W. (1971) Preperceptual auditory images. J. Exp. Psychol. 85, 411-417.

Mattingly, I.G., Liberman, A.M., Syrdal, A.K. and Halwes, T. 1971. Discrimination in speech and nonspeech modes. Cognitive Psychology 2, 131-157.

Milner, B. (1962) Laterality effects in audition. In Interhemispheric Relations and Cerebral Dominance, V.B. Mountcastle, Ed. (Johns Hopkins University Press, Baltimore).

Moray, N. (1959) Attention in dichotic listening: Affective cues and the influence of instructions. Quart. J. Exp. Psychol. 11, 56-60.

Penfield, W. and Roberts L. (1959) Speech and Brain Mechanisms. (Princeton University Press, Princeton).

Porter, R., Shankweiler, D., and Liberman, A.M. (1969) Differential effects of binaural time differences in perception of stop consonants and vowels. Proc. 77th Annual Convention of the Amer. Psychol. Assn.

Shankweiler, D. (1966) Effects of temporal-lobe damage on perception of dichotically presented melodies. J. Comp. Physiol. Psychol. 62, 115-119.

Shankweiler, D. and Studdert-Kennedy, M. (1966) Lateral differences in perception of dichotically presented synthetic consonant-vowel syllables and steady-state vowels. J. Acoust. Soc. Amer. 39, 1256(A).

Shankweiler, D. and Studdert-Kennedy, M. (1967) Identification of consonants and vowels presented to left and right ears. Quart. J. Exp. Psychol. 19, 59-63.

Siegel, S. (1956) Non-parametric Statistics (McGraw-Hill, New York).

Spellacy, F. and Blumstein, S. (in press) The influence of language set on ear preference in phoneme recognition. Cortex.

Spreen, O., Spellacy, F. and Reid, J. (1970) The effect of interstimulus interval and intensity on ear asymmetry for nonverbal stimuli in dichotic listening. Neuropsychologia 8, 245-250.

Studdert-Kennedy, M. and Shankweiler, D. (1970) Hemispheric specialization for speech perception. J. Acoust. Soc. Amer. 48, 579-594.

Studdert-Kennedy, M., Shankweiler, D. and Schulman, S. (1970) Opposed effects of a delayed channel on perception of dichotically and monotically presented CV syllables. J. Acoust. Soc. Amer. 48, 599-602.

Treisman, A.M.. (1960) Contextual cues in selective listening. Quart. J. Exp. Psychol. 12, 242-248.

Wada, J. and Rasmussen, T. (1960) Intracarotid injection of sodium amytal for the lateralization of speech dominance: Experimental and clinical observations. J. Neurosurg. 17, 266, 282.

Weiss, M.S. and House, A.S. (1970) Perception of dichotically presented vowels. J. Acoust. Soc. Amer. 49, 96(A).

PART III: PUBLICATIONS AND REPORTS*

Publications and Manuscripts

Audible Outputs of Reading Machines for the Blind. Franklin S. Cooper, Jane H. Gaitenby, and Ignatius G. Mattingly. Bulletin of Prosthetics Research, BPR 10-15, Spring 1971.

A Voice for the Laboratory Computer. Franklin S. Cooper, T.C. Rand, R.S. Music, and I.G. Mattingly. 1971 IEEE International Convention Digest. (Institute of Electrical and Electronics Engineers, Inc., 1971) 104-105.

Speech Acoustics and Perception. Philip Lieberman. In Communicative Disorders: An Introduction to Speech Pathology, Audiology, and Speech Science, ed. by H. Halpern (New York: Random House, in press).

Letter Confusions and Reversals of Sequence in the Beginning Reader: Implications for Orton's Theory of Developmental Dyslexia. Isabelle Y. Liberman, Donald Shankweiler, Charles Orlando, Katherine S. Harris, and Fredericka B. Berti. To be published in Cortex (1971).

Phonetic Ability and Related Anatomy of the Newborn and Adult Human, Neanderthal Man, and the Chimpanzee. Philip Lieberman, Edmund S. Crelin, and Dennis H. Klatt.

On Learning a New Contrast. Leigh Lisker.

Reports and Oral Papers

Neural Specialization for Speech Perception. Michael Studdert-Kennedy. New York Audiological Society, New York, 16 October 1970.

On Disengaging the Speech Processor. Ruth S. Day. Invited address, Academy of Aphasia, New Orleans, 19 October 1970.

Comments on Some Dichotic Experiments. Michael Studdert-Kennedy. Academy of Aphasia, New Orleans, 19 October 1970.

Colloquium. Ruth S. Day. Department of Psychology, Brown University, October 1970.

---

*Most of the contents of this Report, SR 24, are included in this listing.

Perceptual Competition Between Speech and Nonspeech. Ruth S. Day and James E. Cutting. Paper presented at the 80th annual meeting of the Acoustical Society of America, Houston, 3 November 1970.

Perception of Dichotically Presented Steady-State Vowels as a Function of Interaural Delay. Emily Kirstein. Paper presented at the 80th annual meeting of the Acoustical Society of America, Houston, 3-6 November 1970.

Levels of Processing in Speech Perception. Ruth S. Day and James E. Cutting. Psychonomic Society, San Antonio, 5 November 1970.

On the Evolution of Phontic Ability: Neanderthal Man. Philip Lieberman. New England Linguistic Society, Massachusetts Institute of Technology, 7 November 1970.

EMG Studies of the Larynx. Thomas Gay. Fall Colloquium Department of Communicative Disorders, Northwestern University, 12 November 1970.

Opening Remarks. Franklin S. Cooper. Conference on Sensory and Training Aids for the Deaf, Tidewater Inn, Easton, Maryland, 15 November 1970.

Hemispheric Specialization for Speech Perception. Michael Studdert-Kennedy. Colloquium, City University of New York, Graduate Center, 4 December 1970.

Thesis

Temporal Factors in Perception of Dichotically Presented Stop Consonants and Vowels. Emily F. Kirstein. Ph.D. dissertation, University of Connecticut, Storrs.